Cyberwarfare challenge

National cybersecurity plans should go beyond the cold-war mentality of an arms race and focus more on linking traditional computer security with protections for industrial control systems.

If anyone needs proof that a cyber arms race is in the making, they need look no further than last week's news headlines. In the United States, the Pentagon is expected soon to release a report — or at least an unclassified version of a report — describing how the US government might respond to a cyberattack that causes physical damage. One option, according to a report in *The Wall Street Journal*, might be to forgo the subtlety of a cyberresponse, and drop bombs on a suspected attacker.

Similarly, in an editorial in *The Guardian* last week, UK defence minister Nick Harvey revealed the creation of a cyber operations group that would place cyberwarfare on a footing similar to conventional military operations. "Cyber will be part of a continuum of tools with which to achieve military effect, both defensive and otherwise, and will be an integral part of our armoury," Harvey wrote.

It is now nearly a year since the alert was first raised about Stuxnet, the malicious software, or malware, that targeted Iran's nuclear programme. It is hardly surprising, therefore, that governments are now coming forward with plans for cyberwar. Yet the intensive push for cyberweapons only highlights a glaring gap in openly funded research for cyberdefence. Indeed, Stuxnet, possibly the first truly government-backed cyberweapon, was eventually defused not by military cyber warriors, but by private researchers (see page 142).

Stuxnet is clearly a game-changer, demonstrating an ability to target a cyberattack on a grand scale. The subsequent call for the development of cyberweapons sounds very much like the cold-war push to build ever larger nuclear arsenals — and no doubt claims of cyberweaponry gaps will arise. In the United States, the Department of Homeland Security is waging a campaign to engage anybody working with a computer. Governments, it is clear, will have responsibilities to protect infrastructure and installations, but private companies and individuals will be expected to take increasing responsibility for their own and therefore everybody else's protection.

There is a real and profound public vulnerability that should motivate countries to invest in cybersecurity research: the antiquated state of security for industrial control systems, which makes even the most-developed countries as vulnerable as Iran to a Stuxnet-like attack.

It is not just the power plants or the electricity grid that are at risk. In the United States, for example, the water industry is particularly vulnerable, according to experts. Run by small organizations, highly fragmented and with few resources to invest in security, the nation's water supply is an inviting target for attack. The food industry, not traditionally the focus of public concerns about cybersecurity, is another soft spot. Most people don't realize that programmable logic controllers, the devices targeted by Stuxnet, are used to run the heavily automated food-packaging industry. And the Stuxnet history shows that however isolated in cyberspace the target is, there is still a major threat of a spread to any system controlled by software.

On the research side, Stuxnet highlights several challenges worthy of scientific pursuit and government support. Experts point to the need to go beyond the current generation of signature-based anti-virus software. And scientists bemoan the lack of access to good data — in this case viruses — that are needed to help them conduct research. Then there's the need to bridge the gap between traditional computer security and research into industrial-control-system security.

> ***"Stuxnet is proof that governments can bring together different branches of research."***

Despite those challenges, the US government has no clear research agenda, or even research community, that focuses on cybersecurity.

This problem could be tackled in many ways, but the first step might be to create a national research plan for cybersecurity, appointing a lead agency that would coordinate research, and perhaps even funding centres of excellence at various universities to encourage interdisciplinary research.

Stuxnet is proof that governments can — when they so choose — bring together different branches of research and bridge the gap between computer-security researchers and the industrial-control-system-security community. If governments can do this to create cyberweapons, they should be equally capable of driving research into cyberdefence. ■

Second chances

Leaders must end a run of unmet pledges when they meet to discuss sustainable development.

Next June, world leaders will gather in Rio de Janeiro, Brazil, for Earth Summit 2012, to discuss (again) how to steer the planet towards a more sustainable future. The gathering marks the 20th anniversary of the 1992 Rio Earth Summit, at which heads of state agreed on a set of principles intended to guide sustainable global economic growth — including the precept that environmental protection should be central to development.

Little progress has been made on this in the two decades since the summit. Rather, countries have continued to pursue relatively unrestricted economic development, with limited attempts to minimize environmental impacts. So, will next year's summit do any better?

Those who attend will be forced to confront a string of failures to meet international green goals, including a pledge to stem the loss of biodiversity by 2010 — as agreed under the Convention on Biological Diversity — and to set new binding targets to reduce greenhouse-gas

emissions under the United Nations Framework Convention on Climate Change. The chances of meeting the UN's Millennium Development Goals to halve poverty by 2015 look equally unpromising.

It does not help that the UN has been lacklustre in its preparations for the Rio summit. A panel of senior politicians and business heads has been put to work to draw up a plan (again) for global sustainable growth that will set the agenda for much of the discussion. But the panel was announced only last August, and it is not clear that such an important task can be completed in so little time.

Still, many scientists and environmental economists remain hopeful. Last month, *Nature* joined a group of 17 Nobel laureates in Stockholm as they drew up their own vision of the key challenges to sustainable development. Given the size of the task, the mood was surprisingly upbeat. Central to the proposals that the group came up with was the need (again) to change the mindset of world leaders. Rather than keeping to the traditional view that economic development and environmental conservation sit in opposition, the laureates stressed that continued damage to factors such as biodiversity, soil quality and indigenous people's land rights will increasingly affect economic growth.

There are encouraging signs that, in some places, the necessary change in attitudes is under way. For example, late last month at the Global Energy Partnership in Rome, 23 governments agreed on holistic indicators to assess the sustainable production and use of bioenergy. These include the price and supply of food and the net creation of jobs, as well as water quality and greenhouse-gas emissions. The current biofuel fiasco, in which policies on the use of such fuels have been introduced ahead of the proper checks and balances, could have been avoided had these wider factors been given proper consideration.

Similarly, the Organisation for Economic Co-operation and Development in Paris has launched the Your Better Life Index, a tool to evaluate livelihoods using indicators that go beyond gross domestic product. They include housing, environment, safety and work–life balance. As the Nobel laureates noted, such welfare indicators are needed to account properly for natural capital and the social aspects of progress in economic decisions.

"Continued damage to factors such as biodiversity will increasingly affect economic growth."

Governance remains a major issue with the implementation of environmental goals. Part of the reason that the 1992 Earth Summit failed to have the hoped-for impact was that no international body was given responsibility to monitor and enforce its decisions. This remains the case, but suggestions on how to change the situation are maturing. Brice Lalonde, coordinator of the Rio summit, told a meeting in Brussels on 25 May that he wants to see the World Trade Organization's environmental remit strengthened, so that it can police any new global agreement. Others would prefer to see a beefed-up UN Environment Programme collaborate more with other relevant UN and international bodies. There could even be a role for the UN Security Council.

Political realities, or what are still viewed as political realities, remain a huge obstacle to sustainable development. But for those willing to listen, the global community now has at least a wider and more thorough understanding of the scale of the environmental problems it faces. This may yet spur political will to ensure that the Rio summit, and wider discussion on the vital decisions that it represents, are not a waste of time (again). ■

Misspent energy

The crisis at Fukushima Daiichi should spark a rethink of nuclear-research programmes.

Three months after a triple meltdown at the Fukushima Daiichi nuclear power station in Japan, the world is taking a hard look at nuclear power. Last week, Germany said that it would close all of its 17 nuclear plants. Switzerland has also announced its withdrawal from the nuclear arena. Other nations remain committed for now but, in the West, hopes for a nuclear renaissance seem moribund.

Nevertheless, global energy needs continue to rise. By 2020, the world's electricity demand will have increased by 35–40%, according to the International Energy Agency (IEA) in Paris. Without nuclear power, many nations will struggle to meet that demand, especially if they cut back on fossil fuels to reduce the effects of climate change.

Germany hopes to make up its shortfall through an increase in renewable energy sources and a 10% reduction in energy consumption. New technologies could help to meet these targets. Yet in 2009, the latest date for which IEA figures are available, Germany spent US$246 million on nuclear research — roughly what it did on research into renewables and energy efficiency combined. In Japan, which continues to be plagued by energy shortages from Fukushima's shutdown, the US$2.7-billion nuclear-research budget was six times the energy-efficiency budget.

Many of these research programmes began in the 1950s and 60s, when fission reactors seemed to be the first step on the road to a nuclear future. Technical challenges, the enormous expense of fission power and the risks associated with meltdowns have made that road seem much longer today than it did 50 years ago.

Some nuclear investments seem more questionable following the Fukushima crisis and potential gaps in energy provision. Should Japan spend hundreds of millions of dollars on advanced breeder reactors when its plans for conventional ones are on hold? Should Germany continue its sizeable national programme in nuclear fusion, a distant and difficult technology, when its fission reactors are being shut down?

Meanwhile, the threat from climate change grows ever larger, and there is a pressing need for research to help reduce it. More efficient building design could drastically reduce energy consumption, and materials research is needed to drive down the cost of solar panels. New technologies must be developed and integrated into a more robust electricity grid if renewable power is to be efficiently distributed.

However, in many nations, the research is under-supported. A 2010 IEA analysis found shortfalls in all energy research except fission. Even a small shift from nuclear to other areas could make a big difference.

None of this means that nations should abandon fission. Existing nuclear plants continue to provide cheap, carbon-free energy, and some nations, notably China, have decided that, despite safety concerns, nuclear fission is key to expanding their economies while reducing carbon emissions.

Nor does it mean that all nuclear research should be abandoned. Indeed, some of it seems prescient, given the recent disaster: research into nuclear waste disposal will undoubtedly inform the ongoing clean-up at Fukushima (see page 135). And research into conventional light-water reactors could lead to safety improvements. Other endeavours, such as reactors that can produce medical isotopes, stand on their own merits.

But conventional fission is a mature technology. Today's reactor designs are safer and more efficient than those from the Fukushima Daiichi era. They are the ones that countries will build. More advanced reactor designs may be necessary one day, but for now they seem a very expensive dream. Cheaper areas of research could have a bigger impact in the short term. In a world with finite resources, and serious energy and environmental crises on the way, it is time to rethink research priorities. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

WORLD VIEW *A personal take on events*

Beware of gifts that come at too great a cost

Danger lurks for state universities when philanthropy encroaches on academic independence, warns **Sheldon Krimsky**.

America's public universities risk compromising their autonomy and better judgement when, faced with major budget deficits from declining taxpayer revenue, they grasp at opportunities to land external funding from private donors. The financial landscape makes institutions vulnerable to ideological predators who, under the cloak of philanthropy, wish to take control of what is taught and by whom.

The issue has been highlighted by the recent controversy over the 2008 decision by Florida State University (FSU) in Tallahassee to accept US$1.5 million from the Charles G. Koch Charitable Foundation in Arlington, Virginia. Like many public universities, FSU has found it harder to attract high-level faculty members in a financial landscape dominated by state budget cuts, an economic downturn that has hit endowments, and limits placed on tuition fees.

The Koch foundation is an example of private philanthropy with an ideology. Its billionaire founder, Charles Koch, is an advocate of minimalist government (a vestige of a nineteenth-century-style free-market economic system), personal responsibility in lieu of social safety nets, privately financed education, and an end to the government-run social-security system. Koch and his brother David have been among the leading funders of the libertarian Tea Party and support its organizations and political candidates.

Let's be clear. It is not unusual for private donors to support university faculty positions in certain fields. But the FSU case is remarkable for the strings that came attached to the money. I have examined many such agreements, but the one that FSU signed with the Koch foundation breaks troubling new ground.

First, and most publicly discussed, the university agreed to give the foundation the authority to decide the selection criteria used to fill the economics faculty positions that it paid for, and the right to veto candidates of whom it did not approve. This agreement is a marked departure from the well-established separation between private academic philanthropy and faculty hiring decisions.

The university insists that it was aware of the threat to its independence, and was prepared to pull out of the agreement if it felt that its integrity was being undermined by outside influence. It says that the two academics subsequently appointed and funded under the agreement were its choices. Yet it accepts that the contract presents the appearance of outside influence, and says it is now reviewing its decision to sign it.

In my view, the university was at the very least naive, and at most it turned a blind eye to a compromising agreement. FSU should tear the deal up and hand back the cash. This is no idle academic exercise, and there are more problems with the deal than who gets to decide who is hired.

➲ NATURE.COM
Discuss this article online at:
go.nature.com/k9jafo

The stated objective of the FSU–Koch agreement, of which I have a copy, is "to advance the understanding and practice of those free voluntary processes and principles that promote social progress, human well-being, individual freedom, opportunity and prosperity based on the rule of law, constitutional government, private property, and the laws, regulations, organizations, institutions and social norms upon which they rely".

The phrase of most concern is the "practice of those free voluntary processes and principles". Students of political economics will recognize similar phrasing in the nineteenth-century anarchist writings of Peter Kropotkin and Pierre-Joseph Proudhon. Neither classical anarchists nor radical libertarians have any use for strong central authorities that oversee social-welfare programmes. I see no problem with funding professorships in the study of classical anarchism or twenty-first-century libertarianism, any more than I would with funding a Marxist scholar. But the autonomy of the university is transgressed when the criteria for funding seek to advance the practice of a political ideology.

THE CASE IS REMARKABLE FOR THE STRINGS THAT CAME ATTACHED TO THE MONEY.

According to the agreement, performance objectives for the programme will be reviewed by a three-member advisory board, chosen by the Koch foundation, which will monitor the performance of faculty members and check whether they remain true to the programme's mission. The agreement also states that "Individuals holding the sponsored professorship positions will be treated similarly to all other FSU faculty of similar rank". Really? It is inconceivable that the faculty handbook of FSU or any other state university uses "advancement of the practice" of a political ideology to measure academic success. The agreement also stipulates that an "Undergraduate Political Economy Committee" should be set up in the FSU economics department, with one outside member chosen by the foundation. The purpose of this committee is to shape the undergraduate curriculum to ensure that it meets the goals of the agreement. These conditions are unacceptable at any respectable university.

Let there be no mistake: the controversy over the FSU–Koch agreement is not about the diversity of views on economics at America's universities. It is not even, as the university likes to portray, about whether it hired the staff it wanted to. It is about the wider threat to the independence and autonomy of academic appointments, and the proper boundaries between philanthropy and a university's choices about faculty and curriculum. Compromising these values, even under conditions of financial exigency, will turn a university against itself and corrupt its integral value to society. ■

Sheldon Krimsky *is a professor at Tufts University and author of* Science in the Private Interest *and, with Tania Simoncelli,* Genetic Justice: DNA Databanks, Criminal Investigations, and Civil Liberties.
e-mail: sheldon.krimsky@tufts.edu

PARTICLE PHYSICS

Antihydrogen held captive

Hot on the heels of last year's breakthrough, in which physicists briefly captured antihydrogen atoms, comes a 5,800-fold increase in the length of time that atoms of the elusive substance can be trapped.

Researchers working on the ALPHA experiment at CERN, Europe's particle physics laboratory near Geneva in Switzerland, used a magnetic trap to hold a small number of antihydrogen atoms, which consist of a positron orbiting an antiproton, for as long as 1,000 seconds. The achievement may enable studies of the energy levels of antihydrogen, and allow comparison of the properties of matter with antimatter.

Nature Phys. **doi:10.1038/nphys2025 (2011)**

For a longer story on this research, see go.nature.com/fmv4yc

J.-P. CLATOT/AFP/GETTY

GLACIOLOGY

Glacier modelling can reduce risks

Last year, a large, water-filled pocket in the Tête Rousse Glacier of the French Alps threatened to burst and flood the valley below. On an urgent request by local authorities, scientists quickly developed a model to assess whether draining the cavity would weaken the cavity roof and cause it to collapse. The model accurately predicted that draining would be safe.

Taking into account glacier geometry and the physical properties of ice, Olivier Gagliardini and his colleagues at the Joseph Fourier University in Grenoble, France, quantified a reasonably low risk of stress build-up that would burst the cavity during drainage. Measurements of glacial surface displacements taken during and after the successful drainage (**pictured**) confirmed the accuracy of the model's predictions.

Ice-flow models have become mature enough to be used in mitigating glaciological hazards, the team concludes.

Geophys. Res. Lett. **doi:10.1029/2011GL047536 (2011)**

ANTHROPOLOGY

Dinner date in Madagascar

Humans seem to have reached Madagascar many centuries earlier than previously thought.

Dominique Gommery at the French National Centre for Scientific Research in Paris and his colleagues discovered cut marks on bones from pygmy hippopotamuses found in the Anjohibe Cave in the north-west of the island. Although the researchers could not directly date the bones, other samples from the layer in which they were uncovered dated to around 4,000 years ago.

The oldest cut marks known on the island before this were found on bones from the other end of the island that were estimated to be about 1,500 years younger. The Anjohibe bones suggest an earlier human presence and interaction with local fauna.

C.R. Palevol 10, **271–278 (2011)**

CANCER

Probing for pancreatic cancer

Early-stage pancreatic cancer cells can be illuminated in live mice with infrared light.

Pancreatic cancer is one of the deadliest of human cancers, and is often diagnosed only in the later stages of the disease. Dieter Saur at the Technical University of Munich, Germany, and his colleagues found that pancreatic cancer cells in mice express higher levels of certain cathepsin proteins — which break down other proteins — than healthy cells do. The researchers injected a mouse model of the disease with an agent that emits near-infrared light when activated by these proteins. They then inserted a fluorescence-laser-microscopy probe into the animals to detect the light. The pattern of light emission differed between early- and late-stage cancer cells, and the researchers were even able to identify premalignant cells.

If this probe works in humans, it could be used for early screening of at-risk individuals, say the authors.

Proc. Natl Acad. Sci. USA **doi:10.1073/pnas.1100890108 (2011)**

PHOTONICS

Light threaded through tiny gap

When light passes through an opening smaller than half its wavelength, it tends to spread out or 'diffract' — a feature that is problematic to researchers developing

ever-smaller photonic chips. A tiny device designed by Xiang Zhang's group at the University of California, Berkeley, has overcome this diffraction limit.

The device consists of a specially designed semiconductor strip suspended above a metal surface, with a thin gap in between. In this gap, light is converted to an electron wave that can beat the diffraction limit. The electron wave then re-emits the light at the far end of the strip. This gap allows infrared light to travel at 20 times its wavelength, and visible light at 10 times its wavelength.

The team believes that the device could allow optics to be integrated into nanoelectronic devices.

Nature Commun. doi:10.1038/ncomms1315 (2011)

MATERIALS SCIENCE

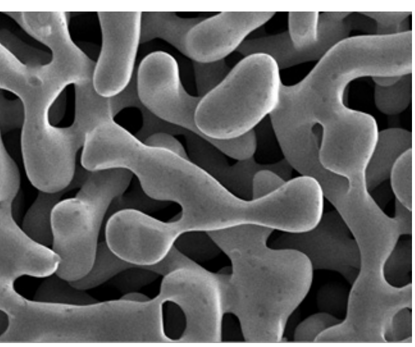
Switching strength on or off

A material has been designed to switch back and forth between a strong, brittle state and a weak, ductile one.

Hai-Jun Jin at the Chinese Academy of Sciences in Shenyang and Jörg Weissmüller at the Technical University of Hamburg in Germany made their composite by imbibing nanoporous gold (**pictured**) with an electrolyte. When the applied electrical potential shifted, the material showed distinct and reversible changes in strength, flow stress and ductility.

Although the exact mechanism remains unknown, the authors think that the adsorbed anions may be slowing down the movement of inherent defects at the surface of the nanoporous gold, enhancing the material's strength. They anticipate the development of a material that can be made ductile during moulding or fabrication, and tuned to the strong state when desired.

SCIENCE/AAAS



GEOSCIENCE

Permafrost thaws, wetlands shrink

As the world warms, wetlands in the Northern Hemisphere may shrink because of thawing permafrost. This finding complicates predictions that wetlands, which harbour methane-producing bacteria, will release more greenhouse gases as temperatures rise.

Permafrost and wetlands are often found together in the Northern Hemisphere. Using a climate model, Christopher Avis at the University of Victoria in British Columbia, Canada, and his colleagues show that, at first, permafrost degradation does increase upper-soil moisture. But as frozen subsurface soil layers continue to melt, water usually trapped near the surface drains to deeper layers, and wetland extent declines.

Nature Geosci. doi:10.1038/ngeo1160 (2011)

IMMUNOLOGY

Cells sense house dust

Household dust may be more allergenic than previously thought — it can activate an innate immune defence.

Anthony Horner at the University of California, San Diego, Mitchell Kronenberg at the La Jolla Institute for Allergy and Immunology in California and their colleagues collected dust samples by vacuuming carpet in various homes. They added the dust to cultures of mouse and human immune cells called invariant natural killer T cells, which quickly recognize and respond to antigens. The researchers found that, in response to the dust, the cells churned out immune-signalling molecules called cytokines.

In a mouse model of allergy, exposing the rodents to house dust and an allergen resulted in more lung inflammation and a stronger immune response than did exposure to the allergen alone.

J. Exp. Med. doi:10.1084/jem.20102229 (2011)

COMMUNITY CHOICE

The most viewed papers in science

CANCER

How cancer cells survive stress

HIGHLY READ on genesdev.cshlp.org 27–31 May

Tumours subject to stressors such as starvation survive by altering their metabolic activities — and they do so by boosting the expression of an enzyme that normally functions in the brain.

Tak Mak at the University of Toronto in Ontario, Canada, and his colleagues found increased expression of *CPT1C* in human lung tumours, and that this elevated expression boosts fatty-acid metabolism and leads to drug resistance. Silencing *CPT1C* renders cancer cells vulnerable to drugs such as rapamycin, low nutrient levels and low oxygen levels.

When human breast cancer cells with silenced *CPT1C* were implanted into mice, they grew at a slower rate than those expressing *CPT1C*. The results suggest that CPT1C mediates tumour-cell survival and could be a useful therapeutic target.

Genes Dev. 25, 1041–1051 (2011)

EVOLUTIONARY BIOLOGY

Stinky, stocky and stripy for a reason

Mammals such as skunks (**pictured**) bear stripes and spots to warn predators of their noxious anal secretions — and these animals have other similarities. Researchers have found that these creatures also tend to be short and stocky, and to live in exposed habitats, suggesting that their bold coloration is a key antipredator defence.

PHOTOLIBRARY

Theodore Stankowich at the University of Massachusetts at Amherst and his colleagues scored dozens of terrestrial mammalian species according to boldness of coloration, body size, habitat openness and ability to use anal secretions in defence.

They found that lineages of bolder, more contrasting coloration are better able to aim and spray their anal secretions at predators.

The authors suggest that the bold colours and enhanced defences evolved as a result of either the mammals moving to a more open environment, which has fewer hiding places, or an increase in predation in their initial habitats.

Evolution doi:10.1111/j.1558-5646.2011.01334.x (2011)

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

Sharing climate data

The World Meteorological Organization (WMO) will create a global framework to give developing countries access to data on local weather and climate, and to help them to prepare for the expected impacts of climate change. The initiative, proposed two years ago (see *Nature* **461,** 159; 2009) was endorsed on 3 June by delegates representing 189 nations at a quadrennial congress of the WMO's governing body in Geneva, Switzerland. It has been given an annual budget of US$75 million.

Nuclear audit

Japan has "underestimated" the tsunami hazard for several of its nuclear plants, according to a fact-finding mission by the International Atomic Energy Agency (IAEA) in Vienna. A preliminary report issued on 1 June recommended that further defences be added to protect plants in the future. But the team praised the Japanese response on the Fukushima nuclear site as "exemplary". On 6 June, Japan's Nuclear and Industrial Safety Agency more than doubled its estimate for the amount of atmospheric radiation released in the first week of the Fukushima disaster, to 0.77 exabecquerels. See page 135 for more.

D. SAMPAIO/AP/PRESS ASSOCIATION IMAGES

Amazon's giant dam approved

Brazil's environment agency, IBAMA, has granted a building licence for the controversial Belo Monte Dam, a massive 11.2-gigawatt hydroelectric plant on the Xingu River — a tributary of the Amazon in the state of Pará. The project's environmental impact had been subject to "robust analysis", IBAMA said. Billed as the world's third-largest hydroelectric plant, the dam will divert water from the river's 'Big Bend' (pictured). It was first planned three decades ago and has been opposed by environmental campaigners and indigenous peoples.

Natural Britain

The first national assessment of the value of Britain's ecosystems, commissioned by the government, was released on 1 June. Building on efforts by the Economics of Ecosystems and Biodiversity international initiative — which makes an economic case for conservation and sustainable use of biodiversity — the study, claim its authors, sets global standards for environmental valuation. It found that yields of most crops and woodland species diversity have improved over the past 60 years. But many more UK 'ecosystem services' (the benefits people get from the natural environment) declined, including fish catches and soil quality. See go.nature.com/r2as7a for more.

BUSINESS

Patent dispute

The US Supreme Court has ruled against Stanford University in California, in a case that tested the boundaries of legislation allowing universities to own intellectual property on federally funded discoveries. In a decision issued on 6 June, the court determined that pharmaceutical firm Roche, of Basel, Switzerland, shares the rights to an HIV test developed by a Stanford University researcher while he was working at Cetus, a company that has since been purchased by Roche. See go.nature.com/kqahzx for more.

RESEARCH

Phones and cancer

Mobile-phone use has joined the purgatorial category of "possibly carcinogenic for humans" as determined by the World Health Organization (WHO). A committee of experts brought together by the International Agency for Research on Cancer, a WHO scientific centre in Lyons, France, announced on 31 May that it could not rule out the possibility that heavy mobile-phone use might increase the risk of brain cancer. The WHO's 'possible carcinogen' category includes 266 other radiation sources and chemicals, including some pesticides, petrol and coffee. Further studies monitoring mobile phones and health are planned. See go.nature.com/hnspnw for more.

New elements

More than a decade after they were first observed, elements 114 and 116 have officially been added to the periodic table. Chemistry's governing body, the International Union of Pure and Applied Chemistry (IUPAC) in Triangle Park, North Carolina, confirmed the elements' discovery in a technical report published on 1 June (R. C. Barber *et al. Pure Appl. Chem.* doi:10.1351/PAC-REP-10-05-01; 2011). Scientists at the Joint Institute

for Nuclear Research in Dubna, Russia, and the Lawrence Livermore National Laboratory in California, created the elements, which still don't have names. Evidence for elements 113, 115 and 118 has not met the criteria for discovery.

Halting melanoma

Two drugs have been shown to slow the progression of advanced melanoma, a particularly lethal form of skin cancer. Vemurafenib will get a priority review by the US Food and Drug Administration after achieving 84% survival rates (after six months) in a large clinical trial. And the recently approved melanoma drug ipilimumab improves average survival by two months, another large trial showed. Its maker, Bristol-Myers Squibb, will unite with vemurafenib's maker, Roche, to test the two drugs in combination. The trial results were presented on 5 June at the American Society of Clinical Oncology's annual meeting in Chicago, Illinois. See go.nature.com/zdmgu3 for more.

EVENTS

Shuttle retired

NASA's space shuttle *Endeavour* landed safely back on Earth on 1 June (**pictured**), completing its last spaceflight and the penultimate mission

T. FARRAR & T. GRAY/NASA

of the shuttle fleet. In 25 space missions, *Endeavour* travelled nearly 200 million kilometres and orbited Earth some 4,671 times. It will now be exhibited at the California Science Center in Los Angeles.

E. coli outbreak

The source of the outbreak of enterohaemorrhagic *Escherichia coli* sweeping across northern Europe had still not been clearly identified when *Nature* went to press. By 7 June, 23 people had died and more than 2,400 had been infected. The outbreak has sparked calls for changes to disease-surveillance systems. See page 137 for more.

FUNDING

Cheaper vaccines

Pharmaceutical companies have agreed to large cuts in the prices of the vaccines that they sell to the GAVI Alliance — a global health partnership based in Geneva, Switzerland, that focuses on getting vaccines into low-income countries. The cuts, from firms that include GlaxoSmithKline, Merck, Johnson & Johnson and Sanofi, cover vaccines against rotavirus and human papillomavirus, and a combined vaccine against five lethal diseases. Despite the savings, GAVI still needs US$3.7 billion to enable a planned $6.8-billion expansion of vaccination programmes in 2011–15. A key meeting for donors' pledges will be held in London on 13 June. See go.nature.com/b2sagg for more.

People power

The Wellcome Trust — Britain's largest non-governmental funder of biomedical research — has named the first researchers to benefit from a new scheme that directs money to people rather than projects (see *Nature* **462,** 145; 2009). Twenty-seven scientists will receive £56 million (US$92 million), with individual awards ranging from around £1 million to £3 million.

Genomics retreat

The Jackson Laboratory, a medical research centre based in Bar Harbor, Maine, has scrapped plans to set up a major satellite facility for personal genomics research in Florida. The centre had spent more than a year negotiating with two districts for a site, but announced that it was pulling out on 3 June, after Florida politicians determined that the state could not afford to invest the US$100-million co-funding that Jackson wanted.

COMING UP

15–17 JUNE
Stanford University in California hosts the Fifth International Meeting on Synthetic Biology, which includes discussions of the field's future.
http://sb5.biobricks.org

15–18 JUNE
The International Society for Stem Cell Research holds its annual meeting, in Toronto, Canada.
www.isscr.org/meetings

AWARDS

Winning millions

The three 2011 Shaw prizes, worth US$1 million apiece, were announced on 7 June. Astronomy prizewinners are Enrico Costa of Rome's Institute of Space Astrophysics and Cosmic Physics, and Gerald Fishman at NASA's Marshall Space Flight Center in Huntsville, Alabama. They led space missions demonstrating the origin of γ-ray bursts. The life sciences and medicine prize went to Jules Hoffmann of the University of Strasbourg, France, Ruslan Medzhitov at Yale University in New Haven, Connecticut, and Bruce Beutler at the Scripps Research Institute in La Jolla, California, for discovering the mechanisms of innate immunity. Demetrios Christodoulou at ETH Zurich in Switzerland and Richard Hamilton at Columbia University in New York won the mathematics prize; their work on nonlinear partial differential equations has applications in general relativity.

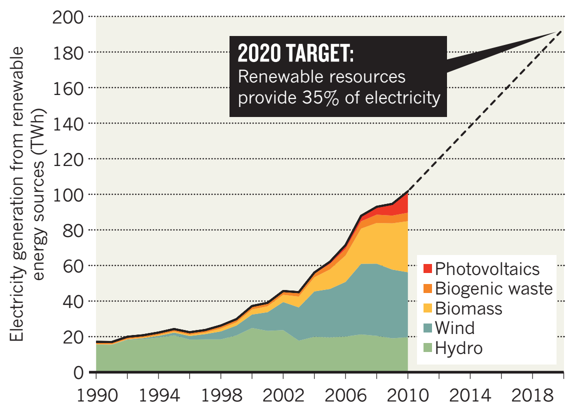
TREND WATCH

Germany's government wants to double the country's use of renewable resources to generate electricity by 2020. While cutting overall demand, it would increase renewables from 17% (or 100 terawatt-hours a year) of the total today to 35%. If the country maintains its current pace (see chart), it will meet that target. But even this doubling would not fill the energy gap left by the government's decision to shut down all the country's nuclear power stations by 2022. See go.nature.com/wordrf for more.

SOURCE: BMU

GERMANY'S RENEWABLES SURGE

If it maintains its recent pace, Germany can meet domestic targets to double electricity generation from renewables by 2020.



➲ NATURE.COM
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

G. WEBB/IAEA

On a visit to Fukushima in late May, safety experts praised Japan's response to the nuclear disaster.

NUCLEAR ENERGY

Fukushima deep in hot water

Rising levels of radioactive liquid hamper clean-up effort.

BY GEOFF BRUMFIEL AND DAVID CYRANOSKI

In the first hours of March's accident at the Fukushima Daiichi nuclear power plant, workers rushed to flood three damaged reactors with sea water to prevent a catastrophic meltdown. Three months later, water is still being pumped into the cores and has become the biggest obstacle to cleaning up the site.

Buildings there are deep in radioactive water, slowing work to a crawl. Storage tanks are rapidly reaching capacity and, if the trend continues, drainage trenches will start overflowing as early as 20 June, according to a report last week from the plant's operator, the Tokyo Electric Power Company (TEPCO).

TEPCO plans to install a decontamination system, which it hopes will remove radioisotopes from the water so that it can be reused to cool the reactors. The system should slow the consumption of water and reduce the danger posed by the wastewater. But even when the system is up and running — expected to be on 15 June — it will generate large amounts of radioactive waste, leading some critics to question whether it is the best solution.

Residual nuclear decay in the three reactors — which all suffered total meltdowns — means that they will need cooling for many months to come. TEPCO switched to using fresh water two weeks after the accident, because the salty sea water they had been using was extremely corrosive to the stainless-steel reactor vessels. The water in the reactors remains salty, even though several thousand litres of fresh water are being poured into the reactors every hour.

More than 100,000 tonnes of water are now swilling around various parts of the site. In early April, TEPCO was forced to dump more than 10,000 tonnes of low-level contaminated water into the Pacific Ocean, and it has since admitted that several hundred tonnes of highly contaminated water also leaked out, exposing marine life to large doses of radiation (see *Nature* doi:10.1038/news.2011.326; 2011). TEPCO says it has now stemmed those leaks.

The radioactive water is hampering work to bring the reactors under control and prepare for their decommissioning, says Jack DeVine, a retired nuclear consultant who led water-decontamination efforts after the 1979 Three Mile Island nuclear meltdown in Pennsylvania. Basements filled with radioactive water also put workers on the site at risk. "It makes even small things nightmarish to do," he says.

Rather than simply pouring ever more water into the hot reactors, TEPCO had planned to establish a closed-loop cooling system that would recirculate water through the cores. But extensive damage to the reactors has made that impossible.

Under a revised plan posted on the company's website on 17 May, the reactors will continue to be filled with a steady flow of fresh water, which will then be pumped out of the basements of the reactor buildings, decontaminated and circulated through the cores again. TEPCO estimates that it will need to decontaminate some 250,000 tonnes of water by mid-January 2012, when it hopes the reactors will finally be cool enough to shut down permanently. But the system will be far less efficient than the original plan, and will ultimately increase the amount of waste to be handled.

The decontamination system is being built by Paris-based nuclear manufacturer Areva and nuclear-remediation company Kurion, headquartered in Irvine, California. The water will pass through Kurion's filters, which contain a zeolite mineral — an extremely porous aluminosilicate that loosely binds metal ions. Through a combination of adsorption and ion exchange, the filters will trap the radioactive elements strontium-90, caesium-134 and caesium-137, reducing their concentration in the water by a thousand times.

Areva's process will then take over. The ▶

water will pass into a series of tanks, where it will mix with reagents such as nickel ferrocyanide and barium sulphate, along with polymers and sand. The dissolved radioactive metals will form precipitates and colloids, which can be trapped as a radioactive sludge, allowing the water to be desalinated and fed back into the reactors. The two processes should reduce the concentration of caesium — the major element of concern from the reactors — in the water by up to a million times. TEPCO estimates that the process will cost ¥53.1 billion (US$660 million). Areva says it has successfully tested its system with sea water containing radioisotopes and borates — a simulation of the chemical brew in the reactors, where boric acid was added to halt nuclear reactions.

But some experts in Japan have expressed reservations about the decontamination process. Radioactive water will continue to flow from the cores into basements and trenches, and damage to the site means there will probably be further leaks. Ming Zhang, who studies environmental pollution risks at the National Institute of Advanced Industrial Science and Technology in Tsukuba, fears that contaminated water will end up in the ocean.

Kenji Takeshita, a specialist in water treatment at the Tokyo Institute of Technology, says that although a zeolite filtration system worked at Three Mile Island, the water pumped through it was fresh. “This time the water is full of salt,” he says. The chemical similarity between sodium and caesium ions may make the zeolite extraction process far less efficient, he says.

Areva’s system will also generate up to 2,000 cubic metres of hot, radioactive sludge by next January. Compared with solid waste, which can be encased in cement for long-term storage, the sludge will need more elaborate containment to prevent it from leaking out into the environment. TEPCO says that the Areva system was added to “ensure the success of the extraction system”, and is now drafting plans for dealing with the zeolite and sludge waste.

Even when the water is dealt with, Japan will face a bigger contamination problem. Radioactive soil has been removed from around schools and other institutions near Fukushima, but it currently sits in large mounds or has been buried at shallow depths. Chihiro Inoue, an expert in soil and groundwater remediation at Tohoku University, estimates that hundreds of thousands of tonnes of radioactive soil will have to be dealt with. “They’re not even thinking about what to do with that yet,” he says. ■ SEE EDITORIAL P.128

➲ WWW.NATURE.COM/JAPANQUAKE

Q&A Yusuke Nakamura
A healthier Japan

Yusuke Nakamura, a geneticist at the University of Tokyo, has long argued that the country should reform its health-care system and work harder to cash in on its biological discoveries. His mission was given fresh urgency by the Tohoku earthquake and tsunami in March — and now he has a pulpit. In January, Nakamura launched the Office of Medical Innovation, a cabinet-level advisory organization, and he will make his case at a high-level government meeting next week.

MARIKO IKEDA

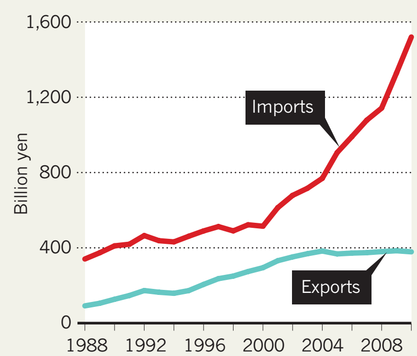
How did the disaster expose weaknesses in Japan’s medical system?
Many people lost their medical records, so the government in Tokyo didn’t know the volume of drugs needed in different places. And medical support is terrible. If you have a stroke in Soma, you’ll have to drive an hour and a half or two hours to the hospital in Sendai or elsewhere. It’s an urgent situation.

Victim identification has been delayed because the government is trying to do it on the basis of dental records, which often aren’t available. Two thousand recovered bodies remain unidentified. I argue that they should use single-nucleotide-polymorphism testing, which is faster and more accurate. I estimate it would cost ¥800 million (US$10 million) — nothing compared with the recovery costs.

How could these problems be avoided?
In Japan, each hospital has its own medical-record system. We need instead to have a national ‘cloud’ system to store records, and patients could carry their own records on a data card or a phone. We should start with Tohoku, where the system needs reconstructing anyway, and then extend to the rest of the country. More destructive quakes will occur.

DRUG IMBALANCE

The gap between Japan's pharmaceutical imports and exports has grown sharply over the past decade.



SOURCE: MINISTRY OF FINANCE, JAPAN

With a good system in place, we could save a large number of lives.

What health studies are you advocating to monitor tsunami survivors?
We have proposed three kinds of cohort study that would cover a total of about 500,000 people. The first would monitor the psychological condition of survivors. A study in Thailand [after the 2004 Indian Ocean tsunami] found that one-fifth of people had post-traumatic stress syndrome. That would mean a huge number in Japan. We need to start monitoring, so that we can treat it at an earlier stage.

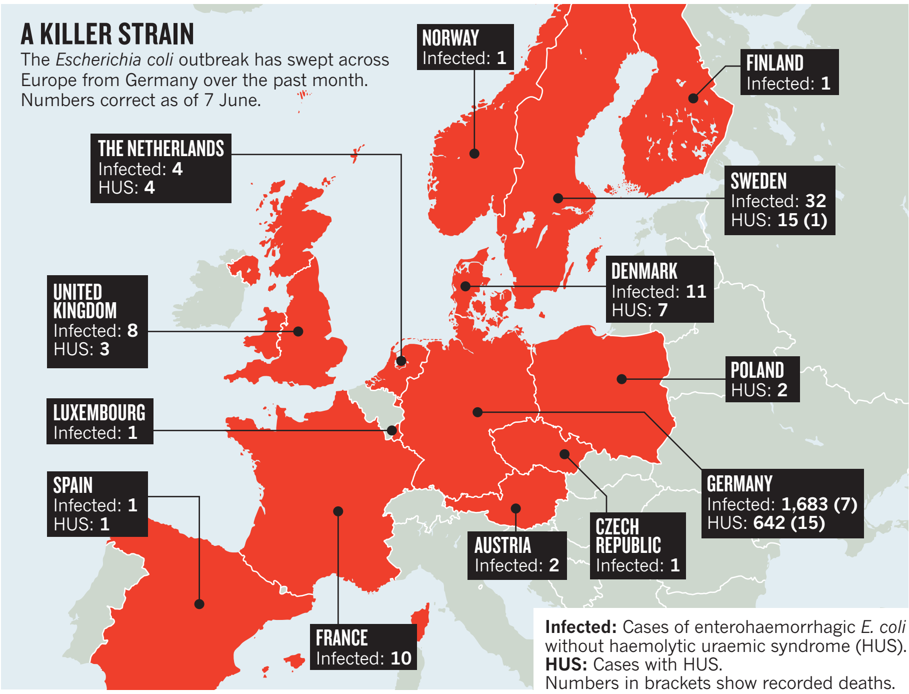
The second study would look at health impacts and chronic disease in those who lived through the disaster and were evacuated for extended periods. We would also have to consider the effects of persistent low-level radiation [from the damaged Fukushima nuclear plant]. The third cohort would be for workers and residents exposed to high doses of radiation. That would have to go on for 20–30 years.

How good is medical innovation in Japan?
It’s terrible. We import over $10 billion more than we export in drugs (see ‘Drug imbalance’). Japanese scientists have papers in *Nature*, *Science* and *Cell*, but few publish in *The Lancet* or *The New England Journal of Medicine*, and some say that we need more support for clinical trials. That’s true, but the drugs in the trials would still come mainly from the United States or Europe. What we need is support for applied research. We are going to propose a super-high-throughput screening and drug-optimization facility. There is one in Taiwan that can screen a million compounds per month. We are proposing something on that scale, with an operating cost of $30 million–40 million a year.

We have also proposed a bioinformatics centre that could help to manage the steep increase in genetic and epigenetic data. Integrating them with clinical records could create a personalized medical system within the next ten years.

The government has put a lot of money into life sciences over the past decade, but the ultimate goal was medical innovation. We need to learn why we have failed to achieve that. ■

INTERVIEW BY DAVID CYRANOSKI



SOURCE: EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

MICROBIOLOGY

Microbe outbreak panics Europe

Spread of rare Escherichia coli *strain raises questions over surveillance of infectious diseases.*

BY MARIAN TURNER IN MUNICH

Confronted with what has become one of the world's most severe outbreaks of *Escherichia coli*, physicians and scientists in Germany say that the country's fractured health-management system has failed to handle the crisis properly. They are calling for major reforms so that outbreaks are reported sooner and more modern technology is used to help identify their source, in order to bring health emergencies under control more quickly.

During the past month, a strain of enterohaemorrhagic *E. coli* (EHEC) has infected more than 2,400 people in 13 countries across Europe, killing 23 (see 'A killer strain'). Public-health experts — scattered across many state and federal ministries for health, agriculture and consumer protection — are still trying to pin down where the bacterium came from and why it causes such severe symptoms.

Hospitals recorded the first cases on 1 May, according to the Robert Koch Institute (RKI), the German federal agency for disease surveillance in Berlin. Yet it was not until 22 May that the first report of an unusual number of EHEC infections in Germany arrived at the European Centre for Disease Prevention and Control in Stockholm. This was unusually long — it typically takes 14 days to detect an outbreak, says Angelika Fruth from the RKI.

Several factors conspired to cause the delay. EHEC infections are not common in adults — so physicians might have initially diagnosed a *Salmonella* or viral infection. The microbe also behaves differently to typical EHEC strains when cultured for diagnosis, which hampered scientists trying to identify it. And under the German health system, local authorities only report such infections weekly to state governments — which then have another week to tell the RKI. It was not until 25 May that the rare *E. coli* strain O104:H4 was named as the culprit.

The strain produces proteins that help it stick to food, and to the human gut. This might explain why symptoms seem more severe than in previous *E. coli* outbreaks, including bloody diarrhoea and, in almost one-third of patients, haemolytic uraemic syndrome (HUS), which can cause kidney failure, neurological complications and death. Indeed, so many patients have developed HUS that German physicians have treated more than 200 people with the antibody eculizumab, which had previously only been used to treat three infection-related cases of the syndrome.

Despite all the obstacles, the outbreak could have been identified earlier, says Flemming Scheutz, head of the World Health Organization Collaborative Centre for Reference and Research on Escherichia and Klebsiella in Copenhagen. The polymerase chain reaction (PCR) could have been used to amplify the microbe's genes, he says, enabling scientists to identify it in hours rather than the days it took to culture and test bacteria from patients' stool samples. "PCR testing for bacteria is already in routine use in livestock and food, and hospitals already have the technical platforms to do PCR," says Scheutz, "but these tests just aren't used regularly in people."

Fruth agrees that a PCR protocol would be highly desirable. However, even though it is as cheap as culture-based techniques, it is not funded as a standard test for cases of diarrhoea by the German health-care system, she notes.

Faster reporting would also help next time there is an outbreak. The US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, introduced a rapid-surveillance platform called PulseNet following an outbreak of a different EHEC strain, called O157:H7, in 1993. The system rapidly compares the molecular fingerprints of suspected bacterial infections in food or humans with those of known pathogens. The results are uploaded to a national database, which allows real-time tracking of the infection clusters in an outbreak.

Epidemiologist Christopher Braden from the CDC says that the number of recognized US outbreaks has now doubled thanks to PulseNet, often leading to earlier identification of problems in food production. Europe does not have a similar real-time tracking system, although Denis Coulombier of the European Centre for Disease Prevention and Control says that the existing European surveillance system has worked reasonably well during this outbreak. Moreover, PulseNet might have overlooked the unusual O104:H4 strain, because there is no definitive list of the *E. coli* strains that represent a threat to human health. "We need a constellation of genes, including those for strain markers and toxins, that could be used to identify dangerous bacteria," says Glenn Morris, director of the Emerging Pathogens Institute at the University of Florida in Gainesville.

Comparing the genome of O104:H4 with that of other pathogenic *E. coli* strains should reveal similarities and differences that will be useful in diagnosing future infections. Alexander Kekulé, a microbiologist at the Martin Luther University of Halle-Wittenburg in Halle, Germany, hopes politicians will learn from the outbreak too, and establish a long-overdue national authority for controlling outbreaks. ■

FUNDING

US research facilities anticipate budget gloom

Interdisciplinary centres are at the sharp end of proposed cuts.

BY EUGENIE SAMUEL REICH

NATIONAL HIGH MAGNETIC FIELD LABORATORY

The National High Magnetic Field Lab is one of many interdisciplinary centres slated for substantial cuts.

As US federal agencies prepare for lean times ahead, concern is growing at some of the centralized, multimillion-dollar facilities supported by the National Science Foundation (NSF). These provide users with access to specialized instruments that include powerful X-ray sources, high magnetic fields and large telescopes. Several such centres are expecting cuts in the 2012 budget.

"If the proposed budget goes through, the impact on users will be quite severe," says Sol Gruner, director of Cornell's High Energy Synchrotron Source (CHESS), which the NSF has told to expect a possible 25% cut. The NSF-funded source of high-energy X-rays is used by around 1,000 researchers a year to characterize materials from semiconductors and superconductors through to proteins and even artwork.

Congress and the White House remain at odds over the 2012 federal budget, but the political furore surrounding the budget deficit almost guarantees that most government departments and agencies will face some austerity measures. The NSF's 2012 budget request to Congress calls for a decrease in the proportion of its budget going into research infrastructure, from 24% in the provisional 2011 budget to 22% in 2012 (see 'Scaling back'). Given an overall increase in the budget request, this amounts to an increase of 3.9% in infrastructure investment.

But some facilities may suffer more than the figures published so far suggest because of the discontinuation of funds awarded under the stimulus bill, the American Recovery and Reinvestment Act (ARRA), and other supplemental grants. The ARRA doled out $361 million to facilities in 2010 — on top of a facilities budget of $954 million.

On paper, for example, CHESS and the Cornell Electron Storage Ring look to be due a 71.9% increase in funding in 2012. But Gruner says this does not seem to take account of supplementary grants, some of which were distributed under ARRA, that he has been told will not continue. The result is the predicted 25% cut to the current budget of $20.3 million. Because of fixed staff costs, Gruner says, this could reduce the light source's running time by as much as 50%.

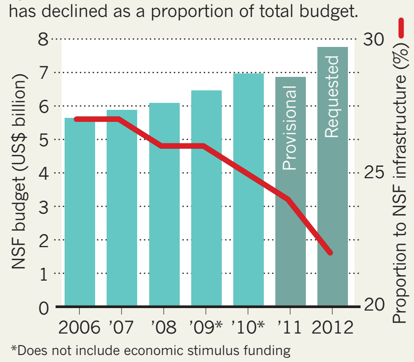
➲ NATURE.COM
For more US science budget stories, visit:
go.nature.com/eg2noo

Other centres also face reductions. The National Nanotechnology Infrastructure Network (NNIN), a group of 14 centres with specialized nanotechnology equipment, is expecting a 4.3% cut. At one NNIN centre, the Cornell NanoScale Science and Technology Facility, this is not enough to reduce running time, but is likely to drive an increase in user fees, says Dan Ralph, the centre's director. The National High Magnetic Field Laboratory (NHMFL) at Florida State University in Tallahassee, used by scientists studying quantum effects in substances ranging from graphene to biological molecules, is facing a 6.4% cut.

Greg Boebinger, the director of the NHMFL, says that the interdisciplinary nature of the users of some NSF-funded centres puts them at risk of disproportionate cuts. "The challenge is always to find interdisciplinary funding for interdisciplinary research," he says.

SCALING BACK

Infrastructure expenditure, including facilities, by the US National Science Foundation (NSF) has declined as a proportion of total budget.



*Does not include economic stimulus funding

Meanwhile, Ian Robertson, director of the NSF's Division of Materials Research (DMR), which funds interdisciplinary facilities including CHESS and the NHMFL, acknowledges that in planning for fiscal year 2012 the division's facilities budget was cut by 15%, even though the research budget is up by 12.8%. But he denies that the NSF's priorities have changed. "There is no shift away from facilities," he says. "We have a complex portfolio and we have to balance our roles."

Two astronomical facilities, the National Optical Astronomy Observatory, which runs telescopes on Kitt Peak in Arizona, and the National Astronomy and Ionosphere Center, which runs the Arecibo Observatory in Puerto Rico, face cuts of 7.4% and 17.9%, respectively, under the fiscal year 2012 budget.

Both the DMR and the NSF's astronomical sciences division will hold external peer reviews later this year that are expected to address the balance between facilities and research grants.

Ray Bowen, a mechanical engineer at Texas A&M University in College Station who chairs the NSF's oversight board, says that there has been no specific advice from the board to pull out of facilities, but it is possible that NSF management has decided to do so as one way to trim its budget. "That would be a prudent decision," he says. ■

FACILITIES

Hungarian natural history under threat

Historical collections given marching orders as government plans military university at museum site.

BY MARIAN TURNER

Looking for a new home: 200 human mummies from the eighteenth century, the remains of rare European dinosaurs and 10 million other artefacts currently at the Hungarian Natural History Museum, which is facing eviction later this year. The Hungarian government plans to turn the historic Budapest building given to the museum after the fall of communism in 1989 into a university to train the military or the police.

Scientists in Hungary and abroad are shocked by the move because the imposing 1836 Ludovika building has been extensively renovated for the museum, and curators are still moving the collections in. They say that the museum has not been offered an alternative site, and fear that the collections will have to be stored in crates until a new home is found.

"When the government announced the new university in February, they described the Ludovika as a long-neglected building. That came as a surprise to those of us who work there," says József Pálfy, a member of a joint research team between the museum's palaeontology research group and the Hungarian Academy of Sciences. But the government justifies its decision by saying that parts of the building need further renovations and that using the Ludovika for the new university is in keeping with tradition — the building contained a military academy until 1945.

The museum employs more than 70 scientists and publishes around 50 papers a year in international journals. In addition to the mummies — which were found in a church crypt in Vac in central Hungary and used to study the history of tuberculosis — it houses fossils found in western Hungary from ceratopsian dinosaurs, which were previously thought not to have lived in Europe.

N. BAUER, HNHM

Museum's mummies seek new resting place.

The collections, some of which date back to 1802, had been scattered around the city before the museum was granted the Ludovika buildings in the early 1990s. The buildings were in disrepair, but the Hungarian government invested around 10 billion forints (US$53 million) to refurbish them. The buildings now give the museum 5,000 square metres of exhibition space, as well as modern research laboratories and three underground levels for storage.

András Jávor, state secretary for the Hungarian Ministry of National Resources, which is responsible for the museum, says that no jobs or resources will be lost in the reorganization, and that his ministry "is consulting with the museum about its future location". But Attila Ősi, a palaeontologist in the same research group as Pálfy, whose discovery of the ceratopsian dinosaur fossils led to a *Nature* paper last year (A. Ősi *et al. Nature* **465,** 466–468; 2010), says that research will suffer if they are forced to pack up their specimens again.

About 100 international researchers use the collections every year, and those contacted by *Nature* echo the concerns of their Hungarian colleagues. "The collections at the museum are unique, and moving them again would create huge problems for multinational research collaborations," says Gareth Dyke, a palaeontologist at University College Dublin in Ireland, who is currently working at the museum.

Museum staff had just started to get comfortable at the Ludovika. "The scientists here are still spending time checking inventories to make sure all the objects have survived moving in," says Ősi. "After 200 years we got a central building for our museum," adds István Matskási, its director-general, "and now we do not know where we will have to go." ■

ETHICS

Whistle-blower claims his accusations cost him his job

University denies it retaliated against researcher who questioned supervisor's data.

BY EUGENIE SAMUEL REICH

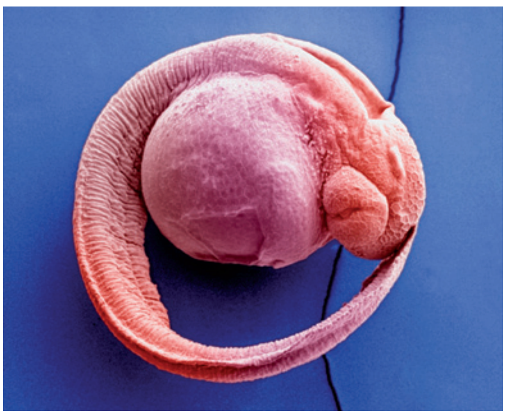
After months of friction that culminated in his openly questioning the reproducibility of data published by his supervisor, a postdoc at the University of Wisconsin–Madison's zoology department was presented with three options. The department's chairman said he could wait to be fired, resign voluntarily or accept a "gracious exit strategy" that would give him time to prepare a paper for publication, if he dropped his "scientific misconduct issues".

When geneticist Aaron Taylor objected that the third option sounded like a "plea bargain" meant to discourage him from pressing his concerns about the lab's data, the chairman, Jeffrey Hardin disagreed. But Hardin also said: "I think you'd have to decide which is more important to you." He later added: "You have to decide whether you want to kind of engage in whistle-blowing."

Taylor recorded the November 2009 exchange without Hardin's knowledge — something permitted under Wisconsin law. Although Taylor declined to make a decision in the meeting, he resigned a few days later, minutes before a disciplinary meeting at which he would have been fired. He says he lost his job because he voiced doubts about data published by the lab of Yevgenya Grinblat, the faculty member who employed him. Grinblat, whose scientific conduct has been upheld by the university and who is not under investigation, says that Taylor lost his job because of a "disrespectful attitude".

The story illustrates a problem that can arise when a junior researcher comes into conflict with a senior colleague over research integrity. Whistle-blowers are supposed to be protected from retaliation, even if they are wrong, as long as they are acting in good faith. Yet, when personality clashes enter the mix — a likely scenario when accusations begin to fly — the outcome can be much less clear cut.

In Taylor's case, Hardin strongly denies any retaliation, and says that a move was made to terminate Taylor's contract for "serious personnel issues". E-mails provided by Grinblat show that she had mentioned and sought advice about what she said was Taylor's negative attitude to colleagues inside and outside the university before he questioned her data. Even so, an e-mail from Hardin to Eric Wilcots, an associate dean for mathematical and natural sciences at the university, seems to imply that Taylor's accusations played a part in his departure. The e-mail said that Taylor had engaged in "seemingly inappropriate correspondence" — referring to a caustic e-mail to Grinblat in which Taylor alleged that she acted improperly as a researcher and a lab supervisor.



Zebrafish embryos are at the centre of a dispute over research integrity.

R. KESSEL/VISUALS UNLIMITED/CORBIS

C. K. Gunsalus, a lawyer and expert on research misconduct at the University of Illinois in Urbana–Champaign, who read documents provided by the university at *Nature*'s request, says: "On the face of it, the university's actions at least raise a question as to whether there was retaliation for speaking out."

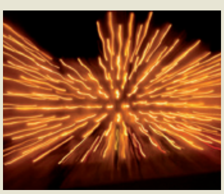
Reprisal against whistle-blowers is banned by the US federal policy on research misconduct, which guides how allegations should be handled by institutions that receive grant money from the government, which Grinblat's lab does. The University of Wisconsin also has a policy that explicitly protects whistle-blowers from retaliation. Taylor says that his experience therefore holds a lesson for those who wish to question a supervisor's work: being explicit about suspicions and citing relevant statutes when making allegations might help to prevent reprisals.

Taylor was studying zebrafish development in Grinblat's lab when he became concerned about work the lab was submitting for publication. In one paper (*Development* **136,** 3791–3800; 2009), on which he was listed as a co-author, Taylor felt that images of zebrafish embryos that were being used to illustrate changes in the expression of particular proteins had been chosen without quantitative support. E-mail records show that he queried the data with Grinblat, but they were nevertheless submitted for publication. Grinblat responds that the use of the images followed standard methods and is based on repeatable experiments. She adds that the quantification method Taylor may have wanted to use was not sensitive enough to detect the changes in gene expression they were observing.

Taylor says that he now cannot prove that the images amounted to falsification because he does not have access to the raw data. He requested access under Wisconsin's

MORE ONLINE

TOP STORY



P. M. ROGERS/ALAMY

Moving mirrors make light from nothing go.nature.com/gz2w8p

OTHER NEWS

- Less than 10% of tropical forests are managed sustainably go.nature.com/aepg4o
- Chronic fatigue syndrome: life after XMRV go.nature.com/eoyxk1
- Did some human ancestors evolve in Europe? go.nature.com/ocjv3v

EXPLAINER

F. DOTT/GREENPEACE/AP

The knock-on effects of Germany's nuclear phase-out go.nature.com/wordrf

public-records laws, but the university declined to release them, in part because it determined that releasing a faculty member's data would violate academic freedom.

Some months after they disagreed on the images, matters came to a head when Grinblat asked Taylor to provide a detailed protocol for a lab technique. Taylor accused her of repeatedly pressuring him to publish data that he considered unreliable, and threatened to bring a case against her to the dean. Soon after, he wrote to Hardin saying that he had doubts about the reproducibility and reliability of work published by the lab.

When asked about Taylor's charges of being pressured, Grinblat says she gave him a deadline because she felt that was the only way to get him to publish his good data. Grinblat comments that, in general, Taylor was extraordinarily reluctant to publish his work.

DIFFERENCE OF OPINIONS

Taylor's former PhD supervisor, neuroscientist Justin Fallon at Brown University in Providence, Rhode Island, says that Taylor is a smart and original thinker who worked independently in his lab and published his work expeditiously. "He worked in a professional and efficient manner," Fallon says, adding that he doesn't agree with the suggestion that Taylor is difficult in the lab.

A letter to Taylor provided by the university says that Wilcots reviewed the evidence that Taylor submitted and talked with him and others before deciding that his complaints did not meet the university's criteria for research misconduct. "What he was presenting was a difference of opinion with his co-author on the significance of results," Wilcots says. However, Wilcots says that he did not look at the raw data from the zebrafish experiments before making that decision. Nor did the university hold an investigation into the matter.

"You have to decide whether you want to kind of engage in whistle-blowing."

Gunsalus says that her interpretation of the documentation is that Taylor's concerns included allegations of misconduct, and would warrant an inquiry.

Because the university asserted that he was not fired, Taylor was not entitled to unemployment benefits in Wisconsin. He therefore filed a complaint with the state's Department of Workforce Development, which ruled that, counter to the university's statements, his departure was not voluntary. In 2010, he also filed a complaint of whistle-blower retaliation with the department, but dropped the case after he took a job at an institution in another state.

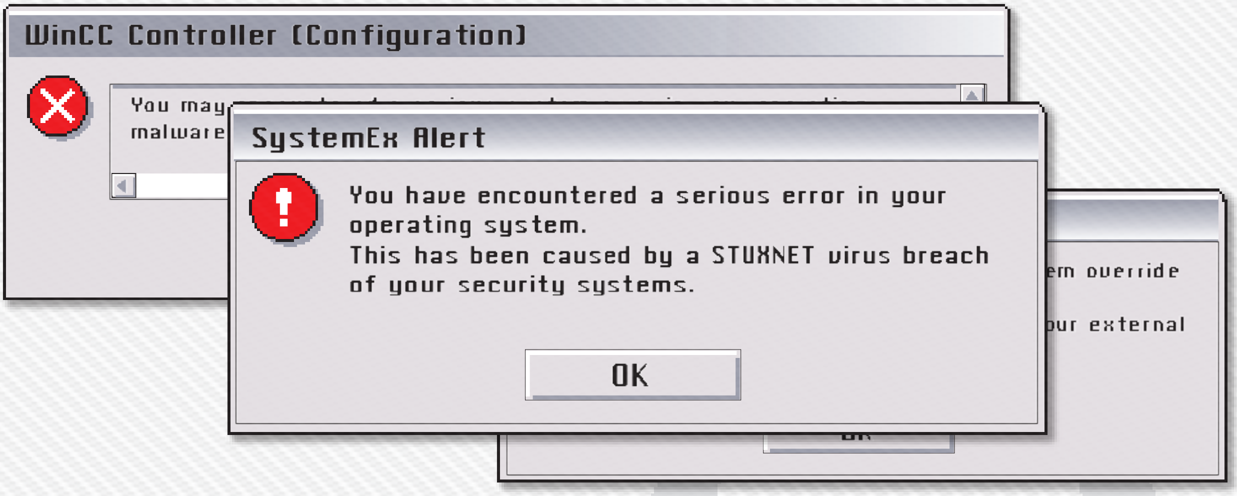
Taylor subsequently forwarded his charges and concerns to the Office of Research Integrity (ORI), which oversees allegations of misconduct in research funded by the National Institutes of Health. The ORI told Taylor in 2010 that it was looking into whether the university had investigated adequately, but would not pursue the allegation of retaliation because Taylor had filed a case at the Department of Workforce Development, even though the case was not longer active.

Wisconsin's associate dean for research policy, William Mellon, says that he did receive a phone call from the ORI about the case, but no further inquiry has been made.

The timing of Taylor's termination and the clash of personalities suggest that extra care was needed to avoid an ambiguous outcome. But Mellon expresses surprise at the suggestion that the university did not act appropriately. He says that the university does take a firm stance on misconduct when it is warranted. "But when it doesn't meet the standard, there are no actions for us to take," he says. Taylor says he has been able to put the situation behind him and is happy at his new institution, which *Nature* agreed not to name, but he remains disillusioned with the way his situation was handled by his former employer.

"It is critical for the integrity of science that universities are not allowed to retaliate against whistle-blowers," he says. ■

Last year's Stuxnet virus attack represented a new kind of threat to critical infrastructure.



IS THIS THE START OF CYBERWARFARE?

by Sharon Weinberger

Just over a year ago, a computer in Iran started repeatedly rebooting itself, seemingly without reason. Suspecting some kind of malicious software (malware), analysts at VirusBlokAda, an antivirus-software company in Minsk, examined the misbehaving machine over the Internet, and soon found that they were right. Disturbingly so: the code they extracted from the Iranian machine proved to be a previously unknown computer virus of unprecedented size and complexity.

On 17 June 2010, VirusBlokAda issued a worldwide alert that set off an international race to track down what came to be known as Stuxnet: the most sophisticated computer malware yet found and the harbinger of a new generation of cyberthreats. Unlike conventional malware, which does its damage only in the virtual world of computers and networks, Stuxnet would turn out to target the software that controls pumps, valves, generators and other industrial machines.

"It was the first time we'd analysed a threat that could cause real-world damage, that could actually cause some machine to break, that might be able to cause an explosion," says Liam O Murchu, chief of security response for the world's largest computer-security firm, Symantec in Mountain View, California.

Stuxnet provided chilling proof that groups or nations could launch a cyberattack against a society's vital infrastructures for water and energy. "We are probably just now entering the era of the cyber arms race," says Mikko Hypponen, chief research officer for F-Secure, an antivirus company based in Helsinki.

Worse yet, the Stuxnet episode has highlighted just how inadequate are society's current defences — and how glaring is the gap in cybersecurity science.

Computer-security firms are competitive in the marketplace, but they generally respond to a threat such as Stuxnet with close collaboration behind the scenes. Soon after VirusBlokAda's alert, for example, Kaspersky Lab in Moscow was working with Microsoft in Redmond, Washington, to hunt

down the vulnerabilities that the virus was exploiting in the Windows operating system. (It was Microsoft that coined the name Stuxnet, after one of the files hidden in its code. Technically, Stuxnet was a 'worm', a type of malware that can operate on its own without needing another program to infect. But even experts often call it a 'virus', which has become the generic term for self-replicating malware.)

One of the most ambitious and comprehensive responses was led by Symantec, which kept O Murchu and his worldwide team of experts working on Stuxnet around the clock for three months. One major centre of operations was Symantec's malware lab in Culver City, California, which operates like the digital equivalent of a top-level biological containment facility. A sign on the door warns visitors to leave computers, USB flash drives and smart phones outside: any electronic device that passes through that door, even by mistake, will stay there. Inside the lab, the team began by dropping Stuxnet into a simulated networking environment so that they could safely watch what it did. The sheer size of the virus was staggering: some 15,000 lines of code, representing an estimated 10,000 person hours in software development. Compared with any other virus ever seen, says O Murchu, "it's a huge amount of code".

Equally striking was the sophistication of that code. Stuxnet took advantage of two digital certificates of authenticity stolen from respected companies, and exploited four different 'zero day vulnerabilities' — previously unidentified security holes in Windows that were wide open for hackers to use.

Then there was the virus's behaviour. "Very quickly we realized that it was doing something very unusual," recalls O Murchu. Most notably, Stuxnet was trying to talk to the programmable logic controllers (PLCs) that are used to direct industrial machinery. Stuxnet was very selective, however: although the virus could spread to almost any machine running Windows, the crucial parts of its executable code would become active only if that machine was also running Siemens Step7, one of the many supervisory control and data acquisition (SCADA) systems used to manage industrial processes.

Many industrial control systems are never connected to the Internet, precisely to protect them from malware and hostile takeover. That led to another aspect of Stuxnet's sophistication. Like most other malware, it could spread over a network. But it could also covertly install itself on a USB drive. So all it would take was one operator unknowingly plugging an infected memory stick into a control-system computer, and the virus could explode into action.

MURKY MOTIVES

It still wasn't clear what Stuxnet was supposed to do to the Siemens software. The Symantec team got a clue when it realized that the virus was gathering information about the host computers it had infected, and sending the data back to servers in Malaysia and Denmark — presumably to give the unknown perpetrators a way to update the Stuxnet virus covertly. Identifying the command and control servers didn't allow Symantec to identify the perpetrators, but they were able to convince the Internet service providers to cut off the perpetrators' access, rerouting the traffic from the infected computers back to Symantec so that they could eavesdrop. By watching where the traffic to the servers was coming from, O Murchu says, "we were able to see that the majority of infections were in Iran" — at least 60% of them. In fact, the infections seemed to have been appearing there in waves since 2009.

The obvious inference was that the virus had deliberately been directed against Iran, for reasons as yet unknown. But the Symantec investigators couldn't go much further by themselves. They were extremely knowledgeable about computers and networking, but like most malware-protection teams, they had little or no expertise in PLCs or SCADA systems. "At some point in their analysis they just couldn't make any more sense out of what the purpose of this thing was, because they were not able to experiment with the virus in such a lab environment," says Ralph Langner, a control-system security consultant in Hamburg, Germany.

Langner independently took it upon himself to fill that gap. Over the summer, he and his team began running Stuxnet in a lab environment equipped with Siemens software and industrial control systems, and watching how the virus interacted with PLCs. "We began to see very strange and funny results immediately, and I mean by that within the first day of our lab experiment," he says.

> "We are probably just now entering the era of the cyber arms race."

Those PLC results allowed Langner to infer that Stuxnet was a directed attack, seeking out specific software and hardware. In mid-September 2010, he announced on his blog that the evidence supported the suspicion that Stuxnet had been deliberately directed against Iran. The most likely target, he then believed, was the Bushehr nuclear power plant.

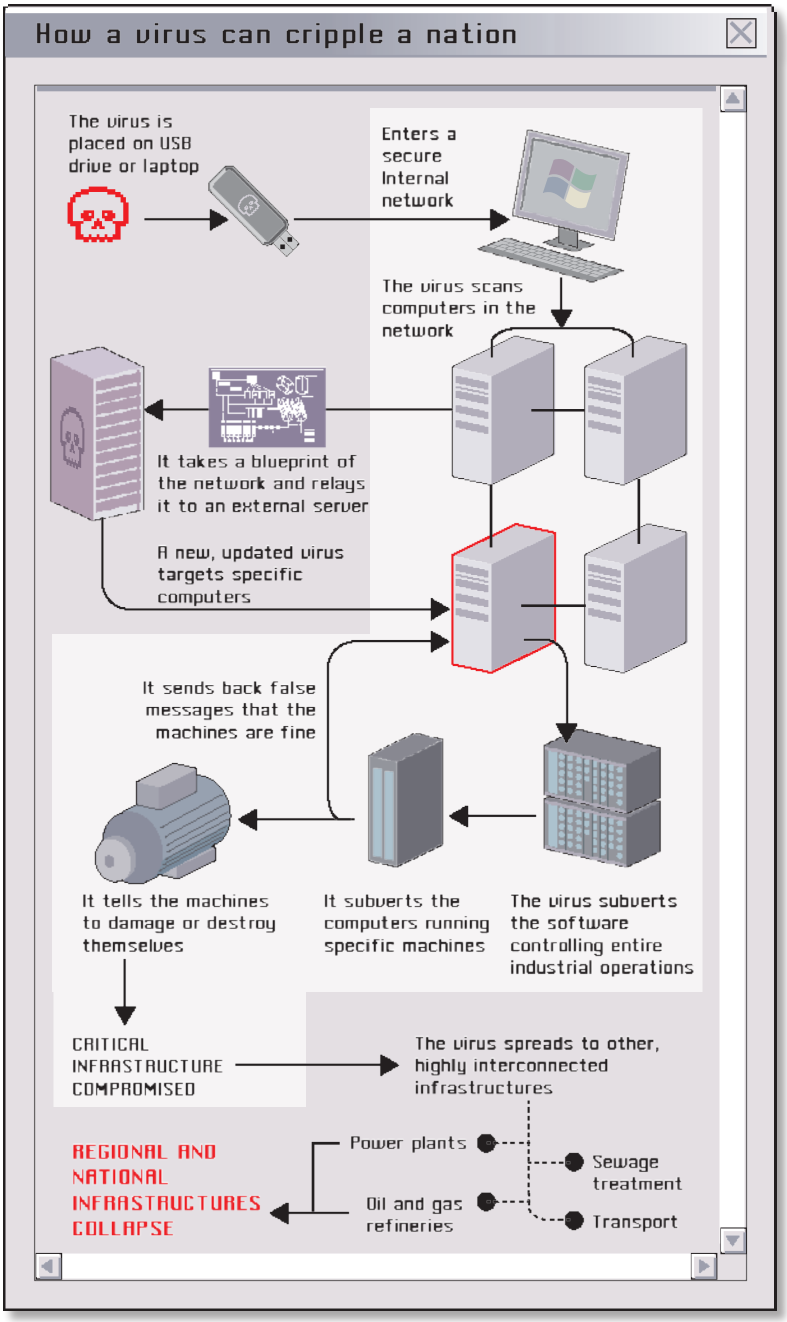
INDUSTRIAL SABOTAGE

Speculative though Langner's statements were, the news media quickly picked up on them and spread the word of a targeted cyberweapon. Over the next few months, however, as Langner and others continued to work with the code, the evidence began to point away from Bushehr and towards a uranium-enrichment facility in Natanz, where thousands of centrifuges were separating the rare but fissionable isotope uranium-235 from the heavier uranium-238. Many Western nations believe that this enrichment effort, which ostensibly provides fuel for nuclear power stations, is actually aimed at producing a nuclear weapon. The malware code, according to Langner and others, was designed to alter the speed of the delicate centrifuges, essentially causing the machines to spin out of control and break.

That interpretation is given credence by reports from the International Atomic Energy Agency (IAEA) in Vienna, which document a precipitous drop in the number of operating centrifuges in 2009, the year that many observers think Stuxnet first infected computers in Iran.

True, the evidence is circumstantial at best. "We don't know what those machines were doing" when they weren't in operation, cautions Ivanka Barszashka, a Bulgarian physicist who studied Iranian centrifuge performance while she was working with the Federation of American Scientists in Washington DC. "We don't know if they were actually broken or if they were just sitting there." Moreover, the Iranian government has officially denied that Stuxnet destroyed large numbers of centrifuges at Natanz, although it does acknowledge that the infection is widespread in the country. And IAEA inspection reports from late 2010 make it clear that any damage was at most a temporary setback: Iran's enrichment capacity is higher than ever.

However, if Natanz was the target, that does suggest an answer to the mystery of who created Stuxnet, and why.



Given the knowledge required — including expertise in malware, industrial security and the specific types and configurations of the industrial equipment being targeted — most Stuxnet investigators concluded early on that the perpetrators were backed by a government.

Governments have tried to sabotage foreign nuclear programmes before, says Olli Heinonen, a senior fellow at the Belfer Center for Science and International Affairs at Harvard University in Cambridge, Massachusetts, and former deputy director-general of the IAEA. In the 1980s and 1990s, for example, Western governments orchestrated a campaign to inject faulty parts into the network that Pakistan used to supply nuclear technology to countries such as Iran and North Korea. Intelligence agencies, including the US Central Intelligence Agency, have also made other attempts to sell flawed nuclear designs to would-be proliferators. "Stuxnet," says Heinonen, "is another way to do the same thing."

Langner argues that the government behind Stuxnet is that of the United States, which has both the required expertise in cyberwarfare and a long-standing goal of thwarting Iran's nuclear ambitions. Throughout the summer of 2010, while Langner, Symantec and all the other investigators were vigorously trading ideas and information about Stuxnet, the US Department of Homeland Security maintained a puzzling silence, even though it operates Computer Emergency Readiness Teams (CERTs) created specifically to address cyberthreats. True, the CERT at the Idaho National Laboratory outside Idaho Falls, which operates one of the world's most sophisticated testbeds for industrial control systems, did issue a series of alerts. But the first, on 20 July 2010, came more than a month after the initial warning from Belarus and contained nothing new. Later alerts followed the same pattern: too little, too late. "A delayed clipping service," said Dale Peterson, founder of Digital Bond, a SCADA security firm in Sunrise, Florida, on his blog.

"There is no way that they could have missed this problem, or that this is all a misunderstanding. That's just not possible," says Langner, who believes that the Idaho lab's anaemic response was deliberate, intended to cover up the fact that Stuxnet had been created there.

But even Langner has to admit that the evidence against the United States is purely circumstantial. (The US government itself will neither confirm nor deny the allegation, as is its practice for any discussion of covert activity.) And the evidence against the other frequently mentioned suspect, Israel, is even more so. Symantec, for example, points out that a name embedded in Stuxnet's code, Myrtus, could be a reference to a biblical story about a planned massacre of Jews in Persia. But other investigators say that such claims are beyond tenuous. "There are no facts" about Israel, declares Jeffrey Carr, founder and chief executive of Taia Global, a cybersecurity consulting company in Tysons Corner, Virginia.

THE AFTERMATH

The 'who?' may never be discovered. Active investigation of Stuxnet effectively came to an end in February 2011, when Symantec posted a final update to its definitive report on the virus, including key details about its execution, lines of attack and spread over time. Microsoft had long since patched the security holes that Stuxnet exploited, and all the antivirus companies had updated their customers' digital immune systems with the ability to recognize and shut down Stuxnet on sight. New infections are now rare — although they do still occur, and it will take years before all the computers with access to Siemens controllers are patched.

If Stuxnet itself has ceased to be a serious threat, however, cybersecurity experts continue to worry about the larger vulnerabilities that it exposed. Stuxnet essentially laid out a blueprint for future attackers to learn from and perhaps improve, say many of the investigators who have studied it. "In a way, you did open the Pandora's box by launching this attack," says Langner of his suspicions about the United States. "And it might turn back to you guys eventually."

Cybersecurity experts are ill-prepared for the threat, in part because they lack ties to the people who understand industrial control systems. "We've got actually two very different worlds that traditionally have not communicated all

that much," says Eric Byres, co-founder and chief technology officer of Tofino Industrial Security in Lantzville, Canada. He applauds Symantec, Langner and others for reaching across that divide. But the effort required to make those connections substantially delayed the investigation.

The divide extends into university computer-science departments, say Byres, himself an ex-academic. Researchers tend to look at industrial-control security as a technical problem, rather than an issue requiring serious scientific attention, he says. So when graduate students express interest in looking at, say, cryptography and industrial controls, they are told that the subject is not mathematically challenging enough for a dissertation project.

"I'm not aware of any academic researchers who have invested significantly in the study of Stuxnet," agrees Andrew Ginter, director of industrial security for the North American group of Waterfall Security Solutions, based in Tel Aviv, Israel. Almost the only researchers doing that kind of work are in industrial or government settings — among them a team at the Idaho National Laboratory working on a next-generation system called Sophia, which tries to protect industrial control systems against Stuxnet-like threats by detecting anomalies in the network.

One barrier for academics working on cybersecurity is access to the malware that they must protect against. That was not such a problem for Stuxnet itself, because its code was posted on the web shortly after it was first identified. But in general, the careful safeguards that Symantec and other companies put in place in secure labs to protect the escape of malware may also inadvertently be a barrier for researchers who need to study them. "If you're doing research into biological agents, it's limited groups that have them and they are largely unwilling to share; the same holds true for malware," says Anup Ghosh, chief scientist at the Center for Secure Information Systems at George Mason University in Fairfax, Virginia. "To advance the field, researchers need access to good data sets," says Ghosh, who was once a programme manager at the US Defense Advanced Research Projects Agency, and is now working on a malware detector designed to identify viruses on the basis of how they behave, rather than on specific patterns in their code, known as signatures.

Academic researchers are also inhibited by a certain squeamishness about digital weaponry, according to Herb Lin, chief scientist at the Computer Science and Telecommunications Board of the US National Research Council in Washington DC. He points out that to understand how to guard against cyberattacks, it may help to know how to commit them. Yet teaching graduate students to write malware is "very controversial", he says. "People say, 'What do you mean: you're training hackers?'"

PREPARING FOR THE NEXT ATTACK

A study last year by the JASON group, which advises the US government on science and technology matters, including defence, found broad challenges for cybersecurity (JASON *Science of Cyber-Security*; MITRE Corporation, 2010). Perhaps most important was its conclusion that the field was "underdeveloped in reporting experimental results, and consequently in the ability to use them".

Roy Maxion, a computer scientist at Carnegie Mellon University in Pittsburgh, Pennsylvania, who briefed JASON, goes further, saying that cybersecurity suffers from a lack of scientific rigour. Medical professionals over the past 200 years transformed themselves from purveyors of leeches to modern scientists with the advent of evidence-based medicine, he notes. "In computer science and in computer security in particular, that train is nowhere in sight."

Computer science has developed largely as a collection of what Maxion calls "clever parlour tricks". For example, at one conference, the leading paper showed how researchers could read computer screens by looking at the reflections off windows and other objects. "From a practical point of view, anyone in a classified meeting would go, 'pooh'," he says. "In places where they don't want you to know [what's on the computer screen], there are no windows. Yet, that was the buzz that year."

Maxion sees an urgent need for computer-science and security curricula to include courses in traditional research methods, such as experimental design and statistics — none of which is currently required. "Why does it matter?" he asks. "Because we don't have a scientific basis for investigating phenomena like Stuxnet, or the kind of defences that would be effective against it."

Also troubling for many of the Stuxnet investigators was the US government's lacklustre response to the virus (assuming that it was not the perpetrator). Stuxnet represents a new generation of cyberweapon that could be turned against US targets, but there is no evidence that the government is making the obvious preparations for such an attack — for example, plans for a coordinated response that pools resources from academia, research institutes and private business.

Other countries seem to be taking the threat more seriously. Some of China's universities and vocational colleges have reportedly forged strong connections with the military to work on cybersecurity, for example. And Israel also seems to be exploiting its computing expertise for national security. A few months before the discovery of Stuxnet, Yuval Elovici, a computer scientist and director of Deutsche Telekom Laboratories at Ben-Gurion University of the Negev in Beersheba, Israel, told *Nature* that he was working closely with the country's Ministry of Defense on cybersecurity. He presciently warned that the next wave of cyberattacks would be aimed at physical infrastructures. "What would happen if there were a code injection into SCADA? What if someone would activate it suddenly?" Elovici asked. He and other experts have been warning for several years now that such an attack on SCADA systems controlling the electricity grid could spark nationwide blackouts, or that the safety systems of power plants could be overridden, causing a shutdown or a serious accident. Similar disruptions could hit water and sewage systems, or even food processing plants.

Such attacks, Elovici warned, are both realistic and underestimated. Asked how bad one would be, Elovici was unequivocal. "I think," he said, "it would be much stronger than the impact of setting several atomic bombs on major cities." ■ SEE EDITORIAL P.127

Sharon Weinberger *is an Alicia Patterson Foundation fellow based in Washington DC.*

DARWIN'S CITY

David Sloan Wilson is using the lens of evolution to understand life in the struggling city of Binghamton, New York. Next, he wants to improve it.

BY EMMA MARRIS

M. LINCOLN

David Sloan Wilson is holding a white ceramic dog dish and making the rounds at the Lost Dog Café in Binghamton, New York. "Just like in church!" the biologist jokes, as he collects crumpled dollar bills on this snowy March afternoon. It is Yappy Hour, a fund-raiser for a project to turn an abandoned dirt-bike track into a dog park. The plan is a contender in Wilson's 'Design Your Own Park Competition', a venture that he says is "richly informed" by evolutionary theory — and one of the many community projects that he is running, co-running or up to his neck in. As with most of Wilson's endeavours these days, the motivation is twofold: to improve the quality of life in Binghamton and to study the city as an evolutionary landscape.

Wilson, who works at the State University of New York in Binghamton, has been a prominent figure in evolutionary biology since the 1970s. Much of his research has focused on the long-standing puzzle of altruism — why organisms sometimes do things for others at a cost to themselves. Altruism lowers an individual's chances of passing its own genetic material on to the next generation, yet persists in organisms from slime moulds to humans. Wilson has championed a controversial idea that natural selection occurs at multiple levels: acting not only on genes and individuals, but also on entire groups. Groups with high prosociality — a suite of cooperative behaviours that includes altruism — often outcompete those that have little social cohesion, so natural selection applies to group behaviours just as it does on individual adaptations[1]. Many contend that group-level selection is not needed to explain altruism, but Wilson believes that it is this process that has made humans a profoundly social species, the bees of the primate order.

Wilson originally built the case for multilevel selection on animal studies and hypothetical models. But eight years ago, he decided to come down from the ivory tower and take a closer look at the struggle for existence all around him. A city — with dozens to hundreds of distinct social groups interacting and competing for resources — seemed to Wilson the ultimate expression of humanity's social nature. If prosociality is important in the biological and cultural evolution of human groups, he reasoned, he should be able to observe it at work in Binghamton, a city of about 47,000 people.

THE URBAN LABORATORY

Differences in prosociality, Wilson thought, should produce measurable outcomes — if not in reproductive success, perhaps in happiness, crime rates, neighbourhood tidiness or even the degree of community feeling expressed in the density of holiday decorations. "I really wanted to see a map of altruism," he says. "I saw it in my mind." And with a frisson of excitement, he realized that his models and experiments offered clues about how to intervene, how to structure real-world groups to favour prosociality. "Now is the implementation phase." Evolutionary theory, Wilson decided, will improve life in Binghamton.

He now spends his days in church basements, government meeting rooms, street corners and scrubby city parks. He is involved in projects to build playgrounds, install urban gardens, reinvent schools, create neighbourhood associations and document the religious life of the city, among others. Wilson thrives on his hectic schedule, but it is hard to measure his success. Publications are sparse, in part because dealing with communities and local government is time-consuming. And the nitty-gritty practical details often swamp the theory; the people with whom he collaborates sometimes have trouble working out what his projects have to do with evolution.

➲ NATURE.COM
To read a Q&A with David Sloan Wilson, visit:
go.nature.com/h1khwl

At the Lost Dog, I ask city planner and frequent collaborator Tarik Abdelazim whether he understands why an academic scientist is taking such a proactive interest in the city. He leans against the bar, glass of wine in hand, and addresses Wilson. "I know you talk about 'prosociality', but how that connects to our good friend Darwin, I don't know."

Wilson built his career in evolutionary theory but largely ignored the struggle for existence all around him.

Fellow biologists are also bemused. According to Wilson's former graduate student Dan O'Brien, now a biologist at the Radcliffe Institute for Advanced Study at Harvard University in Cambridge, Massachusetts, many have reacted to Wilson's work with "a mixture of intrigue and distance". That, says O'Brien, "is because he's not doing biology anymore. He's entered into a sort of evolutionary social sciences." Wilson has acquired the language of community organizing and joined, supported and partially funded a slew of improvement schemes, raising the question of whether he is too close to his research. Has David Sloan Wilson fallen in love with his field site?

Binghamton is hard to love. Established in the early nineteenth century, it has long relied on big industry for its identity and prosperity — early on through the Endicott-Johnson Shoe Company and then through IBM, which was founded in the area. But the manufacturers mostly decamped in the 1990s, and since then the city has taken on an aimless, shabby air. Dollar stores and coin-operated laundries fill the gaps between dilapidated Victorian houses and massive brick-and-stone churches. A Gallup poll in March 2011 listed Binghamton as one of the five US cities least liked by its residents (see go.nature.com/tfdkqi). "It is a town that knows it is badly in need of a revival," says Wilson. Even its motto, 'Restoring the pride', speaks of a city clinging to its past and ashamed of its present.

But as in any city, there is variation: some neighbourhoods are friendlier than others, some are more private and some feel downright dangerous. Wilson started his social experiment by measuring levels of prosociality in each neighbourhood. He and his research group gave surveys to teenagers from the local schools asking how often they helped or were helped by others. The researchers also dropped 200 stamped letters all over the city to see where people would help out a stranger by popping them into the postbox. They even recorded the density of Halloween and Christmas decorations. Wilson's group used some of these data to create maps of the city's prosociality[2]. Neighbourhoods in which people take care of each other appear as peaks, whereas selfishness creates valleys. There isn't a straightforward correlation with any socioeconomic factor, Wilson found. Some of the most cooperative people were from "good, poor neighbourhoods".

So Wilson decided to see whether he could raise up the prosocial valleys by creating conditions in which cooperation becomes a winning

strategy — in effect, hacking the process of cultural evolution (See 'A map of prosociality'). He set about this largely by instituting friendly competitions between groups. His first idea was a park-design project, in which neighbourhoods were invited to compete for park-improvement funds by creating the best plan.

But Wilson soon found out that field experiments in real cities can take on lives of their own: different neighbourhoods couldn't get their acts together on the same schedule, so the competition aspect largely disappeared. Instead, he is now working on turning multiple park ideas into reality. The dog park is one. Another is Sunflower Park, the most advanced project to date, but still a sad, mainly empty lot surrounded by a chain-link fence. Children don't spend a lot of time playing here. Undaunted, Wilson is raising funds and laying plans for a relaxing community space flush with trees and amenities. "In a year," he says, "we will serve you a hot dog from the pavilion."

SCHOOLED FOR SUCCESS

Wilson starts more projects than one man can usually handle. If there is an idea to paint a mural in a blighted area, Wilson wants in. The design can be a competition. And sure, he'll have time to help. "No project should remain unborn," he says. He then lets time constraints weed his overstuffed garden of initiatives.

Right now, education is winning a sizeable share of Wilson's attention. He and his graduate students are working with school administrators in several locations to see whether they can improve student performance with programmes that reward good behaviour and encourage group cooperation. Treats, mostly donated by teachers, range from snacks and school supplies to odd but coveted items such as toiletries from hotels. Wilson insists that this approach is fundamentally evolutionary, pointing out that in 1981, psychologist B. F. Skinner described how reinforcement and punishment 'select' for desired behaviour[3]. Wilson is tweaking the school environment so that it selects for prosociality, and is hoping that the student culture will evolve accordingly.

"People are looking to see how this all unfolds and the kind of success he has with it."

The connection is less clear for Miriam Purdy, the principal of Regents Academy, a school for at-risk teenagers at which Wilson has been advising. When I ask how she sees evolution playing into the incentive programmes, Purdy surprises me and Wilson by saying that she doesn't believe in evolution. But that doesn't bother Wilson. It's no matter, he says, if the principles work.

So far, it is hard to tell whether they do. Rick Kauffman, a graduate student who has spent so much time with students and teachers at Regents that he is a de facto staff member, says that no data have been collected on the incentive programmes because the rules have been adjusted weekly. Instead, he and Wilson are comparing the standardized test scores of Regents students with those of a control group of at-risk kids at the more traditional Binghamton High. A higher percentage of Regents students have taken and passed the state tests administered in January, and scientists and administrators alike are looking forward to the results of tests given later in the year.

Wilson has also been studying Binghamton's religious institutions through an evolutionary lens. He is an atheist, but sees religion as a potentially positive source of community cohesion and a centre of meaning in people's lives. He has written extensively about religion as an adaptation of groups[4], and has been funded by the Templeton Foundation in West Conshohocken, Pennsylvania, to study its effects. At the moment, he is trying to find out which of Binghamton's 100 or so congregations are the most and least likely to flourish in the current cultural environment.

Wilson's trait of interest is the 'openness' of churches. Traditional protestant denominations, of which Wilson is fond, tend towards openness: details of belief and moral codes are individual, arrived at after prayer and discussion. Newer, conservative churches that adhere strictly to the Bible as a literal text would be considered less open.

Wilson would like to understand from an evolutionary perspective why the membership of open churches in Binghamton is currently declining, but 'closed' churches are booming. Perhaps uncertain times create a fearful and socially isolated populace, interested in firm and clear guidance. Or perhaps closed churches uplift their members or focus on group solidarity and recruitment. When people's economic and educational situations are better they may become attracted to more open churches. And Wilson says it is possible that the open churches, by allowing congregants to draw their own conclusions in matters of faith, predispose them to losing faith altogether.

A MAP OF PROSOCIALITY

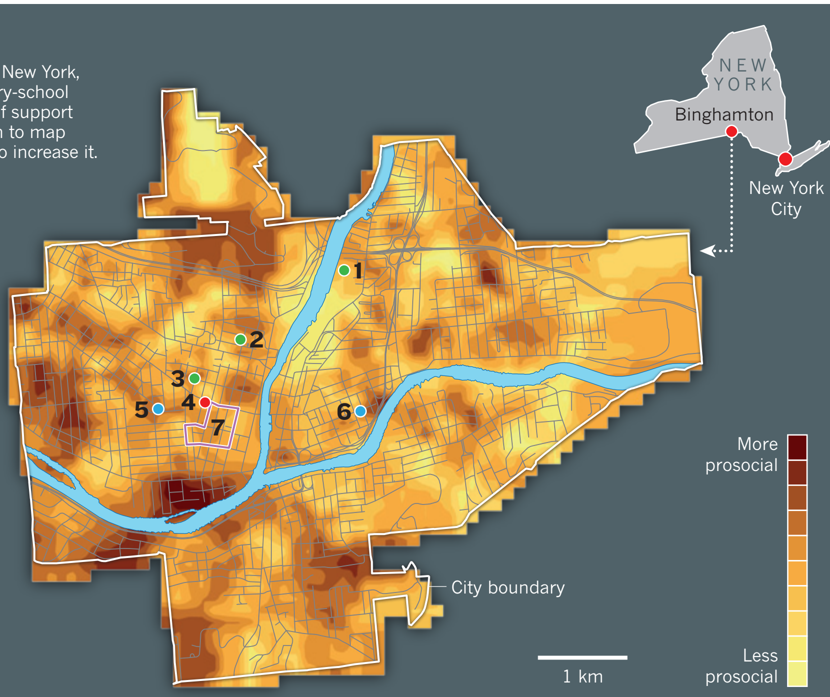
To understand the landscape of social cohesion in Binghamton, New York, David Sloan Wilson and his colleagues surveyed 1,800 secondary-school students on how willing they were to help others, and the kind of support they received from those around them. He used this information to map 'prosociality' throughout the city, then started looking for ways to increase it.

1–3: PARKS
Through the Design Your Own Park competition, community members cooperate to create or improve outdoor spaces, including: **1**, a dog park; **2**, 'Sunflower Park'; and **3**, turning an Edwards Street intersection into a community space.

4: RELIGION
In a pilot study, Wilson and his collaborators have surveyed the congregation at Tabernacle United Methodist Church on matters of prosociality and faith.

5–6: EDUCATION
Wilson has been working with schools, including: **5**, Horace Mann elementary; and **6**, Regents Academy, to see whether incentives to promote cooperation can help to improve student behaviour and test scores.

7: LIVING
Residents of the 'Tree Street District' are being surveyed for attitudes towards different groups before and after the establishment of community-building programmes.



Wilson hopes to test these ideas. The first congregation that he is studying is open, liberal and protestant: the Tabernacle United Methodist Church on Main Street. It has 100 regular members, who barely fill the hulking Victorian building. Last year, Wilson and two graduate students met extensively with the church's Strategic Planning Task Force to craft a survey that was given to everyone on the church's rolls in September.

The survey included questions about prosocial behaviour and about religious faith (such as "How often do you experience the following during worship services at this congregation? A sense of God's presence, Inspiration, Boredom, Awe or mystery, Frustration, Spontaneity, A sense of fulfilling my obligation"). The researchers hope to expand the survey to the roughly 100 congregations across the city, and to create maps describing the ecology of religion in Binghamton.

The work has been met with curiosity and befuddlement. Richard Sosis, an evolutionary anthropologist at the University of Connecticut in Storrs, says that the work is wholly appropriate. "Religion is something that is human, generated from the human mind, which has been designed by natural selection." He adds, "People are looking forward to seeing how this all unfolds, and the kind of success he has with it." However, Ron Numbers, a historian of science and religion at the University of Wisconsin–Madison, is "a little ambivalent and confused". Religious groups develop naturally owing to many historical factors, he says. "Going into some evolutionary account about this doesn't add anything to our knowledge."

There is also the possibility that Wilson has become too close to the church. He has held so many community meetings in Tabernacle's basement that he may even have increased the size of the congregation. I asked Wilson whether the effects he has had on the church will contaminate his study. He pauses for a long time before answering, "Not if you are monitoring what you are doing." Wilson is upfront about his affection for the group. "I'd love this church to grow," he says.

One of Wilson's students on the religion project, Ian MacDonald, says that Wilson has "temporarily" allayed his fears about helping religious organizations. But MacDonald is uneasy about what will happen when they try working with closed, dogmatic churches that condemn homosexuality or teach women to obey their husbands' every command. Wilson says that he is, "sympathetic to the 'niche' occupied by 'closed' churches"; he is not there to judge.

LONG-TERM PROJECT

The playgrounds, schools and churches are but a few of Wilson's ongoing schemes. Another is the West Side Neighborhood Project, which is surveying residents' attitudes towards groups such as black people and students, before and after the creation of a landlord–tenant association and a student association. And he is even doing some work on the genetic level. In Wilson's office in the city, 15-millilitre conical tubes hold DNA samples from elderly Binghamtonites who have been interviewed for their oral histories.

M. LINCOLN

Wilson with Mary Webster, who is coordinating a 'Design Your Own Park' project.

Each of the 47,000 people in Binghamton generates an infinite number of data points, from genomes to attitudes and habits of prayer. Wilson could spend the rest of his life — and those of several graduate students — studying and influencing lives here. But can he study the town and save it at the same time? Ben Kerr, a biologist at the University of Washington in Seattle, says, "In the sense of providing a deeper understanding, evolutionary biology has something to offer." And he likes that Wilson is doing good in the community, but warns, "One has to exercise caution when moving from descriptive to prescriptive stances."

Among social scientists — and that is arguably what Wilson has become — it is not uncommon to become involved with the communities being studied. Such involvement "might be a social good in itself, but it also helps you understand that community better", says Kathleen Blee, head of sociology at the University of Pittsburgh in Pennsylvania. "The desire to promote a social good may be stronger than leaving your field site untouched," she adds.

Still, does one man have time to do both? Mary Webster, a resident who has been working on a park-design project in her neighbourhood, says that she initially saw Wilson as a professor with all the answers. Now, she says she realizes that he is "flying by the seat of his pants". That "sounds about right", says Wilson and, paraphrasing Einstein, he offers, "If we knew what we were doing, it wouldn't be called research."

It is past eight in the evening at Tranquil, one of the few upscale restaurants in town, where Wilson has invited members of the West Side Neighborhood Project. I quiz Wilson's graduate students about how he could possibly pay enough attention to them with all the projects on his plate. But they are loath to say anything negative about the man who is buying them dinner. "He's been better recently," says one.

Meanwhile, Wilson disappears to take a call. A letter had been published in *Nature* that day[5] defending inclusive fitness theory, a framework that seeks to explain the evolution of altruism, and that some evolutionary biologists have called unnecessary. The paper had made ripples in the press, and because Wilson made his name in such theoretical discussions, I assume that he is talking to someone about it. But when he returns, he says that he was on the phone with the religion reporter for the *Binghamton Press & Sun-Bulletin*, giving the scoop on an agreement to survey more congregations.

What seems like big news in the academic world has faded into the background for Wilson. There is so much to do right where he lives. "When you compare what I am doing here with furiously pounding away on my typewriter about that arcane debate, I think I made the right choice," he says. ■

Emma Marris *writes for* Nature *from Columbia, Missouri.*

1. Wilson, D. S. & Wilson E. O. *Q. Rev. Biol.* **82,** 327–348 (2007).
2. Wilson, D. S., O'Brien, D. T. & Sesma, A. *Evol. Hum. Behav.* **30,** 190–200 (2009).
3. Skinner B. F. *Science* **213,** 501–504 (1981).
4. Wilson, D. S. *Darwin's Cathedral* (Univ. Chicago Press, 2002).
5. Abbot, P. *et al. Nature* **471,** E1–E4 (2011).

G. MCDOWELL/NATUREPL.COM

When water is scarce, dust storms strip away the scant soil in Mali.

Save our soils

Researchers must collaborate to manage one of the planet's most precious and threatened resources — for food production and much more, says **Steve Banwart**.

Many researchers focus on how to intensify agriculture for a growing, hungry world. So far, they have largely dodged the question of how global soils will cope.

Our planet's soils are under threat, as witnessed in the past decade by dust-bowl conditions in northwest China, the desertification of grasslands in Inner Mongolia and massive dust storms across north-central Africa. Soil losses in some locations around the world are in excess of 50 tonnes per hectare in a year[1]: up to 100 times faster than the rate of soil formation. In other words, we are losing nearly a half-centimetre layer of this precious resource per year in some places (see graphic).

At the same time, global growth in human population and wealth requires a major intensification of agricultural production to meet an expected 50% increase in demand for food by 2030, and possibly a doubling by 2050[2]. These numbers do not bode well.

Scientists need to develop a predictive framework for soil loss and degradation quickly, to evaluate potential solutions systematically and implement the best ones.

There is a way forward. In the past four years, a global network of research field sites — Critical Zone Observatories — has been established. Multidisciplinary teams are focusing on the fundamentals of soil production and degradation, and aiming to create quantitative, predictive models. This programme has enormous potential. It can and should be accelerated, with stronger collaboration between national programmes and strong links to policy-makers.

CRITICAL LOSSES

Soil lies at the heart of Earth's 'critical zone' — the thin veneer extending from the top of the tree canopy to the bottom of our aquifers. Soil forms as rock breaks up and dissolves, with help from soil organisms, creating particles that bind with decaying biomass and living microbes to form larger aggregates. At least 60% of fertile soil by mass is such 0.25–10-millimetre particles[3]. These aggregates provide a good balance of mineral and organic nutrients, which are processed ▶

by microbes into forms useful for plants. The pores within and between soil aggregates retain sufficient moisture for biological growth, facilitate drainage and allow oxygen to reach plant roots.

The built-up natural capital is lost as soil is washed from fields by rain or snatched away by dust storms during drought. Soil is degraded by pollution, by salts concentrated from evaporating irrigation water, and by compaction by heavy machinery. And it is literally sealed up as cities pave and roof over it — the geographical footprint of Europe's cities has increased by nearly 80% since the 1950s, and continues to expand[4]. In warmer conditions, microbes can degrade organic matter faster, so carbon dioxide and other greenhouse gases are released into the atmosphere, the desired soil aggregates are lost and nutrients are depleted quickly. One study estimates that soils in England and Wales are losing 0.6% of their carbon content each year, probably thanks to climate change[5].

LIFE SUPPORT

Soil does far more than support farming and forestry. It stores carbon, filters water, transforms nutrients and sustains biodiversity[6]. It is not clear how these essential roles will respond to agricultural intensification and other human-driven changes, how they might be enhanced in tandem with farming or how this will affect humanity.

Critical Zone Observatories (CZOs) are designed to help fill these knowledge gaps. The US National Science Foundation (NSF) has invested US$30 million in 6 CZOs and 11 affiliated sites. This is primarily a basic-science endeavour to understand critical-zone development and soil processes in diverse environments — from drylands in Arizona to tropical forests in Puerto Rico, and from coastal Delaware to the Rocky Mountains.

The European Commission funds a $10-million CZO programme with 10 sites in Europe, the United States and China focused on mitigating soil threats. Four core European sites represent key stages of the soil cycle. At the Damma Glacier CZO in Switzerland, researchers study the stages of development of new soil formed over the past 150 years on bedrock exposed as the glacier retreats due to global warming. The Fuchsenbigl CZO in Austria studies the development of soil fertility on a floodplain: sediments deposited along the Danube River since the last glaciation reveal progressive stages of soil formation over thousands of years. The Lysina CZO in the Czech Republic tackles soil recovery in managed forests, in an area damaged by acid rain during the late twentieth century. The Koiliaris CZO in Crete, Greece, has mature soils affected by millennia of agriculture and under imminent threat of desertification because of global warming.

Together, these sites provide data for multiple purposes: to develop a predictive model for soil quality; to put monetary value on soil ecosystem services; and to assess the environmental costs of changing land management. These are essential steps towards proactively designing solutions that could intensify agriculture, halt desertification and store more atmospheric carbon without compromising the future value of soil.

There are now more than 30 CZOs worldwide. They have huge potential to transform soils research. Adding CZOs from the Sahara to the Arctic will help, for example, to assess the impact of climate on soil. This is particularly urgent for the Mediterranean basin, where the Intergovernmental Panel on Climate Change predicts 1–5.5 °C warming this century, mostly in summer, a 30–45% decrease in precipitation and a 6–36% decrease in run-off by 2070[7]. A quantitative model of soil quality could be used to assess the costs and benefits of different mitigation strategies, from enforcing less grazing to revitalizing terracing, encouraging composting or switching to a different mix of crops.

Two key ingredients are still missing from the CZO effort. One is better international integration of research methodology. Work on this is just beginning. In September 2008, the Koiliaris CZO hosted teams from Europe, China and the United States for the first joint training event. In 2010, Pennsylvania State University in University Park ran a field course at the Shale Hills CZO; and the University of Colorado, Boulder, hosts a symposium for young scientists at the Boulder Creek CZO this June. The teams are developing methods to create data sets and models that work across all sites, to study a greater geographical range of critical-zone environments, and to maximize the knowledge being created.

> *"Soil does far more than support farming and forestry."*

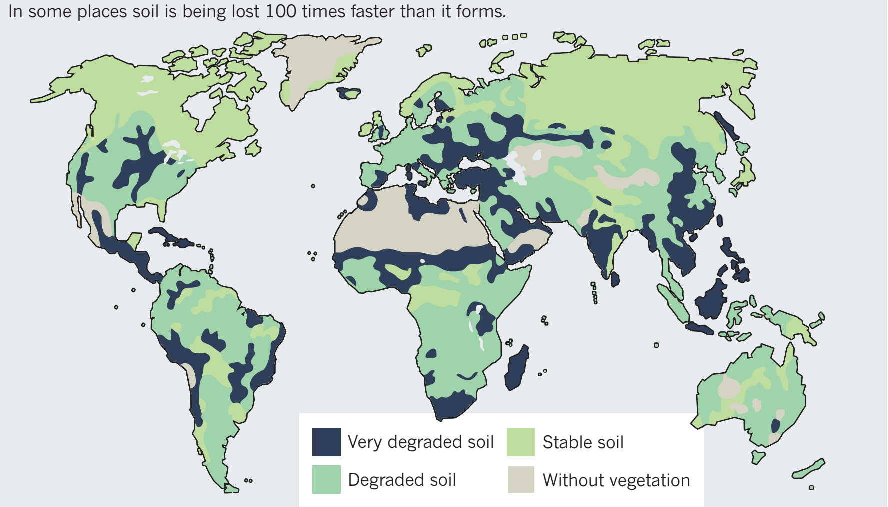
The second missing ingredient is a stronger integration with government and commercial activity. The directed research driven by European policy and the basic research of the NSF programme are a powerful combination. Including commercial partners in research planning will help to translate findings into practical solutions.

The challenge is clear. We need rigorous forecasting methods to quantify and best utilize soil's natural capital, to assess options for maintaining or extending it, and to determine how declines can be reversed. And we need these things well within a decade. ■

Steve Banwart *directs the Kroto Research Institute, University of Sheffield, Sheffield S3 7HQ, UK, and leads the SoilTrEC research consortium. This article was developed in collaboration with members of the SoilTrEC Project Board, www.soiltrec.eu/people.html.*
e-mail: s.a.banwart@sheffield.ac.uk

A THREATENED RESOURCE

In some places soil is being lost 100 times faster than it forms.



SOURCES: P. REKACEWICZ/UNEP/ISRIC/WORLD ATLAS OF DESERTIFICATION 1997

1. Montgomery, D. R. *Proc. Natl Acad. Sci. USA* **104,** 13268–13272 (2007).
2. Godfray, H. C. J. *et al. Science* **327,** 812–818 (2010).
3. Shein, E. V. *Kurs fiziki pochv* (A Course of Soil Physics) [in Russian] (Moscow State Univ. Publishing, 2005).
4. *Urban Sprawl in Europe: The Ignored Challenge.* EEA Rep. 10/2006 (European Environment Agency, 2006); available at http://go.nature.com/uuqmnt
5. Bellamy, P. H., Loveland, P. J., Bradley, R. I., Lark, R. M. & Kirk, G. J. D. *Nature* **437,** 245–248 (2005).
6. European Commission *Thematic Strategy for Soil Protection* COM(2006) 231 (2006).
7. Alcamo, J. *et al.* in *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. & Hanson, C. E.) 541–580 (Cambridge Univ. Press, 2007).

The author declares competing financial interests: for details, and further reading, see go.nature.com/xbjg9j

F. LANTING/NAT. GEOGR.

A forester engages in efforts to eradicate the velvet tree *Miconia calvescens* in Hawaii.

Don't judge species on their origins

Conservationists should assess organisms on environmental impact rather than on whether they are natives, argue **Mark Davis** and 18 other ecologists.

Over the past few decades, 'non-native' species have been vilified for driving beloved 'native' species to extinction and generally polluting 'natural' environments. Intentionally or not, such characterizations have helped to create a pervasive bias against alien species that has been embraced by the public, conservationists, land managers and policy-makers, as well by as many scientists, throughout the world.

Increasingly, the practical value of the native-versus-alien species dichotomy in conservation is declining, and even becoming counterproductive[1]. Yet many conservationists still consider the distinction a core guiding principle[2].

Today's management approaches must recognize that the natural systems of the past are changing forever thanks to drivers such as climate change, nitrogen eutrophication, increased urbanization and other land-use changes. It is time for scientists, land managers and policy-makers to ditch this preoccupation with the native–alien dichotomy and embrace more dynamic and pragmatic approaches to the conservation and management of species — approaches better suited to our fast-changing planet.

The concept of nativeness was first outlined by the English botanist John Henslow in 1835. By the late 1840s, botanists had adapted the terms native and alien from common law to help them distinguish those plants that composed a 'true' British flora from artefacts[3].

Over the next century, many botanists and a few zoologists described and studied introduced species without being aware that others were doing the same. By the time the British ecologist Charles Elton wrote his famous 1958 book *The Ecology of Invasions by Animals and Plants*, some 40 scientists had published descriptions of non-natives, but no consensus had been reached on the desirability of intervening when alien species were introduced.

It wasn't until the 1990s that 'invasion biology' became a discipline in its own right. By this point, partly fuelled by Elton's book, proponents of biodiversity preservation and ecological restoration commonly used military metaphors and exaggerated claims of impending harm to help convey the message that introduced species are the enemies of man and nature.

➲ NATURE.COM
The book that began invasion ecology:
go.nature.com/5aiwqt

Certainly, some species introduced by humans have driven extinctions and undermined important ecological services such as clean water and timber resources. In Hawaii, for instance, avian malaria — probably introduced in the early 1900s when European settlers brought in song and game birds — has killed off more than half of the islands' native bird species. Zebra mussels (*Dreissena polymorpha*), originally native to the lakes of southeast Russia and accidentally introduced to North America in the late 1980s, have cost the US power industry and water utilities hundreds of millions (some say billions) of dollars in damage by clogging water pipes.

But many of the claims driving people's perception that introduced species pose an apocalyptic threat to biodiversity are not backed by data. Take the conclusion made in a 1998 paper[4] that invaders are the second-greatest threat to the survival of threatened or endangered species after habitat destruction. Little of the information used to support this claim involved data, as the original authors were careful to point out. Indeed, recent analyses suggest that invaders do not represent a major extinction threat to most species in most environments — predators and pathogens on islands and in lakes being the main exception[5]. In fact, the introduction of non-native species has almost always increased the number of species in a region[5].

The effects of non-native species may vary with time, and species that are not causing harm now might do so in the future. But the same is true of natives, particularly in rapidly changing environments.

BIOLOGICAL BIAS

Nativeness is not a sign of evolutionary fitness or of a species having positive effects. The insect currently suspected to be killing more trees than any other in North America is the native mountain pine beetle *Dendroctonus ponderosae*. Classifying biota according to their adherence to cultural standards of belonging, citizenship, fair play and morality does not advance our understanding of ecology. Over the past few decades, this perspective has led many conservation and restoration efforts down paths that make little ecological or economic sense.

Take the effort to eradicate the devil's claw plant (*Martynia annua*), introduced from Mexico to Australia in the nineteenth century, probably as a horticultural oddity. For the past 20 years, the Northern Territory Parks and Wildlife Service, along with hundreds of volunteers, have been manually digging up the plants along 60 kilometres of creek bed in Gregory National Park. ▶

K. MOLONEY/THE NEW YORK TIMES/REDUX/EYEVINE; T. & P. LEESON/ARDEA.COM; P. DEL TREDICI; J. WEST/PHOTOLIBRARY

Management of introduced species such as (left to right) tamarisks, pheasants, honeysuckle and zebra mussels should be based on rational, not emotive reasons.

Today, devil's claw is still found in the park and is abundant in adjacent cattle stations. Is the effort worth it? There is little evidence that the species ever warranted such intensive management — it does not substantially change the fundamental character of its environment by, say, reducing biodiversity or altering nutrient cycling[6].

Another example is the US attempt to eradicate tamarisk shrubs (*Tamarix* spp) introduced from Eurasia and Africa into the country's arid lands in the nineteenth century. These drought-, salt- and erosion-resistant plants were initially welcomed into the United States, first as ornamental species for people's gardens and later as shade trees for desert farmers. Then in the 1930s, when water supplies in eastern Arizona, central New Mexico and western Texas ran short, they were indicted as 'water thieves', and later, during the Second World War, as 'alien invaders'. Beginning in 1942, they became the object of a 70-year suppression project involving herbicides, bulldozers and the picturesquely named LeTourneau Tree Crusher[7].

NEW GUIDING PRINCIPLES

Ecologists have since discovered that tamarisks use water at a rate comparable to that of their native counterparts[8]. And the plants are now the preferred nesting habitat of the endangered southwestern willow flycatcher *Empidonax traillii extimus*.

Tamarisks, which survive under common water-management regimes that destroy native trees and shrubs, arguably have a crucial role in the functioning of the human-modified river-bank environment[9]. Yet between 2005 and 2009 alone, the US Congress authorized US$80 million to support ongoing tamarisk control and eradication.

What, then, should replace the native versus non-native species distinction as a guiding principle in conservation and restoration management?

Most human and natural communities now consist both of long-term residents and of new arrivals, and ecosystems are emerging that never existed before. It is impractical to try to restore ecosystems to some 'rightful' historical state. For example, of the 30 planned plant eradication efforts undertaken in the Galapagos Islands since 1996, only 4 have been successful. We must embrace the fact of 'novel ecosystems' and incorporate many alien species into management plans, rather than try to achieve the often impossible goal of eradicating them or drastically reducing their abundance. Indeed, many of the species that people think of as native are actually alien. For instance, in the United States, the ring-necked pheasant, the state bird of South Dakota, is not native to the great plains of North America but was introduced from Asia as a game bird in the latter half of the nineteenth century.

Specifically, policy and management decisions must take into account the positive effects of many invaders. During the 1990s, the US Department of Agriculture (USDA) declared several species of introduced honeysuckles to be alien (harmful), and banned their sale in more than 25 states. Ironically, from the 1960s to the 1980s, the USDA had introduced many of these same species in land reclamation projects, and to improve bird habitats. Recent data suggest that the agency's initial instincts may have been appropriate. In Pennsylvania, more non-native honeysuckles mean more native bird species. Also the seed dispersal of native berry-producing plants is higher in places where non-native honeysuckles are most abundant[10].

Clearly, natural-resource agencies and organizations should base their management plans on sound empirical evidence and not on unfounded claims of harm caused by non-natives. Another valuable step would be for scientists and professionals in conservation to convey to the public that many alien species are useful.

We are not suggesting that conservationists abandon their efforts to mitigate serious problems caused by some introduced species, or that governments should stop trying to prevent potentially harmful species from entering their countries. But we urge conservationists and land managers to organize priorities around whether species are producing benefits or harm to biodiversity, human health, ecological services and economies. Nearly two centuries on from the introduction of the concept of nativeness, it is time for conservationists to focus much more on the functions of species, and much less on where they originated. ■

Mark A. Davis *is DeWitt Wallace professor of biology at Macalester College, St Paul, Minnesota, USA.* **Matthew K. Chew, Richard J. Hobbs, Ariel E. Lugo, John J. Ewel, Geerat J. Vermeij, James H. Brown, Michael L. Rosenzweig, Mark R. Gardener, Scott P. Carroll, Ken Thompson, Steward T. A. Pickett, Juliet C. Stromberg, Peter Del Tredici, Katharine N. Suding, Joan G. Ehrenfeld, J. Philip Grime, Joseph Mascaro, John C. Briggs.**
e-mail: davis@macalester.edu

1. Carroll, S. P. *Evol. Appl.* **4,** 184–199 (2011).
2. Fleishman, E. *et al. Bioscience* **61,** 290–300 (2011).
3. Chew, M. K. & Hamilton, A. L. in *Fifty Years of Invasion Ecology* (ed Richardson, D. M.) 35–47 (Wiley-Blackwell, 2011).
4. Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. *BioScience* **48,** 607–615 (1998).
5. Davis, M. A. *Invasion Biology* (Oxford Univ. Press, 2009).
6. Gardener, M. R., Cordell, S., Anderson, M. & Tunnicliffe, R. D. *Rangeland J.* **32,** 407–417 (2010).
7. Chew, M. K. *J. Hist. Biol.* **42,** 231–266 (2009).
8. Stromberg, J. C., Chew, M. K., Nagler, P. L. & Glenn, E. P. *Rest. Ecol.* **17,** 177–186 (2009).
9. Aukema, J. E. *et al. Bioscience* **60,** 886–897 (2010).
10. Gleditsch, J. M. & Carlo, T. J. *Diversity Distrib.* **17,** 244–253 (2010).

Full author affiliations accompany this article online at go.nature.com/cgbm1y.

A. TANVEER/REUTERS

Three Pakistani men show their scars from selling a kidney — such organ donation was banned in the country in 2007.

ETHICS

Battling the body brokers

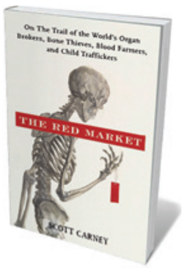
A hard-hitting book calls for greater transparency to deter the illegal trade in human blood, organs and eggs, finds **Laura Spinney**.

You've heard of the black market. Now here's the *The Red Market*, a book describing the global trade in body parts and people that props up the health, fertility and adoption industries. In his exposé of the crimes and human-rights violations that are committed to supply the trade, investigative journalist Scott Carney probes a business that is now "larger, more pervasive, and more profitable than at any other time in history". The World Health Organization estimates that 10% of world organ transplants are obtained illegitimately; Carney argues that it is a good rule of thumb for estimating exchanges of all body parts.

Blood, organs and children have long been traded. Thanks to globalization and advances in technology, so too now are human eggs and surrogate wombs. Carney describes two factors that red markets have in common: first, the transaction is not over when the money changes hands because buyer and seller become biologically linked; and second, because people are squeamish about swapping flesh for money, those buying the human parts tend to describe the exchange euphemistically in terms of altruism — they receive 'donations'.

As there is not enough altruism in the world to supply the insatiable demand for body parts, morally dubious or criminal elements step in. Carney investigates, for example, how the human egg trade in Cyprus exploits poor women from eastern Europe. He also uncovers and carries the distressing news from India to the United States that the adopted son of a

➲ NATURE.COM
For more on medical ethics and clinical trials:
go.nature.com/epaixl



The Red Market: On the Trail of the World's Organ Brokers, Bone Thieves, Blood Farmers, and Child Traffickers
SCOTT CARNEY
William Morrow: 2011. 272 pp. $25.99

midwestern family was stolen from his biological mother while her back was momentarily turned in a Chennai slum.

Other shocking cases include the reported execution of political prisoners in China to provide organs on demand — highlighted in 2006 by United Nations delegate David Matas and retired Canadian politician David Kilgour. Or the blood farm on the India–Nepal border, where a dairy farmer imprisoned poor Nepalese refugees and literally bled them dry, prompting Carney to delve further into India's murky blood trade.

The Red Market is an excellent piece of reporting, but the book is framed around a flawed concept. Carney describes the "specialness" that defines a living person. "There is a clear difference between the living and the dead and that specialness — whatever it may be — is the rock that I've built this book upon," he writes. But there is no such division: death is a process, not an event. Over the centuries, the arbitrary line that society has drawn between life and death has shifted.

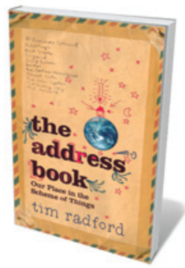
DRAWN OUT DEATH

It used to be that the heart had to stop beating before death could be declared. Now brain death is the usual criterion, and a dead person's heart can continue to beat for a while. As death encroaches on life, more and more people are considered to be eligible for organ harvesting. Carney does not specifically mention this trend, but it inevitably affects organ supply.

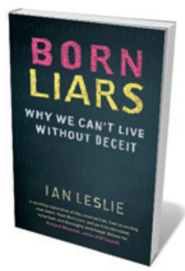
He does, however, allude to the blurring between life and death by describing the trade in a waste product: human hair. Although hair contains DNA, it hardly represents a biological link between donor and wig-wearing recipient, thus failing to satisfy one of his red market factors. It still meets the other criterion: the hair is 'donated', while increasing sums are exchanged for it as it moves up the supply chain.

Temple-goers in Tirupati, India, have their heads shaved for religious reasons, and are not paid for the raw material that ends up in expensive salons in Brooklyn. Likewise, in many countries, women who 'donate' their eggs are prohibited from receiving payment beyond their expenses. Organs too are given away, while every physician, nurse and health coordinator ▶

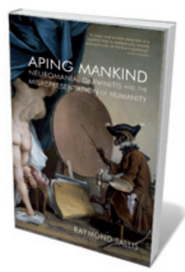
Books in brief



The Address Book: Our Place in the Scheme of Things
Tim Radford FOURTH ESTATE *272 pp. £16.99 (2011)*
As children, many of us will have written down a long version of our address — with our street, home town and country followed by planet Earth, the Solar System, the Universe. In his latest book, science journalist Tim Radford muses on our relationship with locations at increasing scales, asking how they become part of our identities and why we make strong associations with certain places. Starting with his perspective on his possessions, house and town, Radford's horizons expand to encompass the place of humans on the planet, in our Galaxy and in the Universe.



Born Liars: Why We Can't Live without Deceit
Ian Leslie QUERCUS *352 pp. £12.99 (2011)*
Most people say they dislike liars, but we have all told untruths. We may consciously tell a white lie, assuring a friend that an outfit suits them when we know it doesn't. Or we may fool ourselves by adhering to unrealistic beliefs, such as that we will always be healthy so it is fine to smoke. But lies are not necessarily bad, says writer Ian Leslie. Although most people feel psychological discomfort when they tell falsehoods, deceit has positive benefits. As well as playing a part in advertising, politics, sport and war, it is central to human character, he argues, and has evolutionary advantages.



Aping Mankind: Neuromania, Darwinitis and the Misrepresentation of Humanity
Raymond Tallis ACUMEN *416 pp. £25 (2011)*
Human consciousness and behaviour are more complex than can be explained by our brains alone, argues clinical neuroscientist Raymond Tallis in his provocative book. Although he acknowledges that neuroscience has made great strides in recent decades towards understanding how the brain works, he suggests that some scientists have over-reached themselves in believing that everything from art to religious belief can be explained in neural terms. Such 'neuromania' is misleading, he believes.

The Crafting of the 10,000 Things: Knowledge and Technology in Seventeenth-Century China
Dagmar Schäfer UNIVERSITY OF CHICAGO PRESS *352 pp. $45 (2011)*
The demise of China's Ming Dynasty in the seventeenth century was accompanied by a surge in publications detailing the state of knowledge and technology. One of the most significant was Song Yingxing's 1637 volume *Tiangong kaiwu*, or *The Exploitation of the Works of Nature*. It documented the production of materials and goods, from yeast, wine and ink to paper, boats and firearms. In this book, historian Dagmar Schäfer sets Song's encyclopedia within the broader commercial and cultural context of Ming China.

Compassion, by the Pound: The Economics of Farm Animal Welfare
F. Bailey Norwood and Jayson L. Lusk OXFORD UNIVERSITY PRESS *416 pp. $45 (2011)*
Economic assessments have been largely missing from debates on the welfare of farm animals. Agricultural economists Bailey Norwood and Jayson Lusk rectify that in this volume by evaluating the value of organic eggs, free-range pork and the use of antibiotics in farm animals. Consumers are willing to bear the extra costs for greater animal welfare in food production, they find, but the price in some cases is reduced efficiency, which requires more animals to be used.

involved in transplanting them is paid for their contribution.

Because the buyers insist on using the language of gifts, the recompense for the donor — from clinics, hospitals or brokers — is tiny. So only the poor and desperate are tempted. Profiteers increase the supply through coercion.

Protecting donors' anonymity enables buyers to turn a blind eye to the source of their flesh. And as Carney puts it: "The one-two punch of anonymity and donation means that profit-taking middlemen control the entire supply chain."

The global industry in body parts exploits the varied regulations and economic conditions in different countries. The fact that body parts have often travelled across continents also obscures their source. By the time an Indian child's papers reach an adoption agency in the United States, for example, there is often no easy way of verifying whether that child was given up voluntarily.

Banning organ commerce will not help, Carney says, because the red market would be driven further underground and the poorest would continue to supply it in return for risible fees. Nor will fully embracing the market work, as demand will rise alongside supply. Physicians will find new indications for transplants as more organs become available, and as they hold out hope for ever-smaller improvements in the lives of very sick patients. A third option, to grow synthetic body parts, is appealing but remains science fiction for now.

> ***"Every child, kidney or bag of blood should be labelled with the name of the person who provided it."***

Carney does argue that the market should be transparent, with every child, kidney or bag of blood labelled with the name of the person who provided it. The supply pool might shrink, but the criminal middlemen would be eliminated. The richer members of society — those creating most of the demand — could also learn to accept their mortality, and to question whether, in every case, the extended life that a transplant promises is worth the human suffering it costs farther back along the chain. It's not much of a sticking plaster, but *The Red Market* is a reminder that there are some problems that science alone cannot solve. ■

Laura Spinney *is a writer based in Lausanne, Switzerland.*
e-mail: lfspinney@gmail.com

An 1881 *Punch* cartoon satirizing Charles Darwin's body of work that connected humans with worms.

G. BERNARD/SPL

EVOLUTION

Darwin's other books

From geology to mould, the naturalist's publications form a coherent whole, finds **Eugenie Scott**.

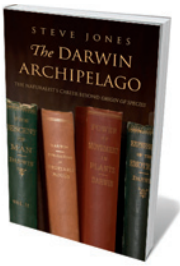
Most people are familiar with *On the Origin of Species*, but few are aware of Charles Darwin's decades of productive work before and after its publication. Indeed, creationists present Darwin as a one-trick pony: a rich dilettante who loafed around his country manor until he stumbled across evolution by natural selection. So I was delighted to read Steve Jones's *The Darwin Archipelago* (published in Britain as *Darwin's Island*), an entertaining and thoughtful treatment of Darwin's other books.

Darwin was a respected scientist — and a recipient of the Wollaston Medal, the Geological Society of London's highest award — well before the *Origin*'s publication in 1859. By then, he had eight books to his name, starting with *The Voyage of the Beagle* in 1839 and followed by three volumes on geology and four on barnacles. Ten further books and monographs appeared after the *Origin*, including *The Formation of Vegetable Mould, through the Action of Worms*, which was published in 1881, the year before his death at the age of 72.

As in his earlier treatment of the *Origin*, titled *Almost Like a Whale* (published in the United States as *Darwin's Ghost*), Jones shows how Darwin's ideas have inspired and been augmented by subsequent research. And Jones shows his flair for a one-liner: a worm is an "animated intestine", and molecular biology is "no more than comparative anatomy plus a mountain of cash".

Jones notes that all of Darwin's work

➲ NATURE.COM
For more, see *Nature*'s Darwin 200 special issue:
nature.com/darwin



The Darwin Archipelago: The Naturalist's Career Beyond Origin of Species
STEVE JONES
Yale University Press: 2011. 248 pp. $27.50

supports the principle that, driven by natural selection, small changes "given time, can produce gigantic ends". Anatomy, behaviour, biogeography, embryology — so many characteristics of organisms can be explained by selection on precursors that gave some advantage to their possessors. The peculiar "contrivances" of insectivorous plants to attract and capture prey, for example, can be explained only by natural selection.

Darwin's works also show what his contemporary William Whewell referred to as "consilience": the confluence of evidence from a variety of sources. As Jones puts it: "The great naturalist's lifelong labours generated an archipelago of information; a set of connected observations that together form a harmonious whole."

Time and again, Jones's book caused me to reflect on how delighted Darwin would have been to have had some titbit of evidence discovered after his death — such as the fossilized bee found with a pollen sac attached, which sheds light on the evolution of plants, or the discovery of genes for olfaction in mice. This latter finding supplements Darwin's observations of the importance of urine marking in mice for choosing a mate and avoiding inbreeding.

The last chapter is depressing but important. Introduced plant and animal species sometimes spread at the expense of natives, reducing diversity, Jones notes. Similarly, our own species has become less diverse: as our ability to manipulate the environment has grown, the variance in human death rates and birth rates has fallen, and variation within and between populations has declined because of migration and intermixing. Yet this variation is the key to adaptation by natural selection.

What does this say for the future? We know that environmental pressures will require adaptation. "One day, [Darwin's] machine will take its revenge," Jones warns. "We may well fail in the struggle for existence against ourselves, the biggest ecological challenge of all." Natural selection bats last. ■

Eugenie Scott *is executive director of the National Center for Science Education, Oakland, California 94609, USA.*
e-mail: scott@ncse.com

Q&A Christiaan Zwanikken
The machinist

Dutch artist Christiaan Zwanikken makes computer-controlled mechanical sculptures, many of which use animal skeletons he has found. As a film about Zwanikken's work and life at an isolated Portuguese convent premieres in Edinburgh, UK, he discusses human relationships with other animals and machines.

Can you describe your artworks?
I take the remains of animals and plants and bring them back to life by means of computer-controlled motors, or servos. The sculptures can interact with people and each other. There is a strong natural element and a technological part; for example, one of my sculptures has two goat skulls that bang into each other using a pneumatic system. Another is a snake eagle that dives down from a height of ten metres to meet a Duracell bunny. There is a hare skull to which I gave a new copper face. And I have five masks with little tongues that repeat lines from the 1982 film *Blade Runner*, in which an android asks for a longer life.

How did you become interested in machines and animals?
My grandfather was an aeroplane technician, and as a child I worked with him in his basement full of exotic parts of aeroplanes. Then my family moved to an isolated convent in rural Portugal and I explored the natural surroundings. I had dreams of machines that could fly, jump and swim. When I was at art school I had three dog skulls in my studio, and I arranged them into a mechanical Cerberus, the many-headed dog that guards the underworld in classical mythology. People were really shocked by it.

What technologies do you use?
I use old-fashioned materials such as cogs and wheels, clockwork devices, electromagnetic coils and servos, as well as radar and ultrasonic sensors. I create an illusion, but also give away my tricks by leaving the mechanisms exposed. Although computer-aided manipulation of materials is getting cheaper, I build my machines by hand. The biggest part of my work is programming them so that they seem to behave naturally.

J. ALTERMAN (CONVENTO)

Convento
DIRECTED BY JARRED ALTERMAN
Edinburgh International Film Festival, 19 June. Sculptures on show 15–26 June.

Paranoia
Gare Saint Sauveur, Lille, France. Until 15 August.

C. ZWANIKKEN, *MONSIEUR HOUDIN, N'OEBLIEZ PAS VOTRE OISEAU!* (2008)/PHOTO: J. ALTERMAN (CONVENTO)

Mechatronics gives an eagle skull new life.

What point are you trying to make?
I am trying to unravel humans' relationship with nature. We have rudimentary reactions to other animals, often a fight-or-flight reaction. There is a dark side to my work, perhaps a warning about the loss of species. But I also try to merge the worlds of animal and machine, so that they coexist peacefully. I try to make something that appears to be real, using technology to create a world that no one has seen before.

What is your most recent piece?
My installation *Scorched Earth*, on show until 15 August at an exhibition in Lille, France, is a post-apocalyptic landscape in which all flora and fauna are extinct and the remains of animals have been combined with machines to populate the world. The artificial animals are autonomous, intelligent and uncontrollable. Humans exist only as a memory. The morning it was finished, I saw the images from Japan after the tsunami. They seemed strikingly familiar. ■

INTERVIEW BY JASCHA HOFFMAN

CORRESPONDENCE

EU innovation must benefit society

The narrow focus on innovation in preparation for the next European Framework programme is steering environmental and health research in the wrong direction (*Nature* **473,** 421; 2011).

Although these fields do not promise immediate economic growth, they are crucial drivers of innovation, focusing on complex socioecological systems, stimulating technological solutions and providing the basis for the sustainability of all innovations.

The European Union's (EU's) 2020 strategy brands the 'Innovation Union' as essential for "smart, sustainable and inclusive growth". This carries the risk that growth, competitiveness and innovation will be taken as ends per se, ahead of human well-being and sustainability. It is worryingly similar to the dominant approach before the economic crisis — one that led to unsustainable use of resources, crippling biodiversity loss and increased greenhouse-gas emissions.

Innovation is not an end, but a means to promote sustainability, human health and well-being. We must target European innovation discourse, policies and actions towards socially meaningful innovation. Technological and ideological lock-ins will not solve economic, societal and environmental crises. Sustainability requires transformation in all spheres (see go.nature.com/rmd44g).

This week in Brussels, the European Commission holds its conference on the Common Strategic Framework for EU Research and Innovation Funding. The European research policy is a powerful tool for shaping "an economy with a human purpose" (R. Passet *Le Monde Diplomatique*, February 2001). The EU should not pass up this historic opportunity.

Sybille van den Hove *Median, Barcelona, Spain.*
sybille@median-web.eu
Jacqueline McGlade *European Environment Agency, Copenhagen, Denmark.*
Michael H. Depledge *Peninsula Medical School, University of Exeter, UK.*

NIH: translation centre bridges a gap

We disagree with Henry Miller that the US National Institutes of Health (NIH) should abandon plans to establish NCATS, the National Center for Advancing Translational Sciences (*Nature* **472,** 169; 2011).

Researchers in academia all too often lack access to drug-discovery infrastructure, hindering translation of NIH-funded basic research into new medicines. Although the pharmaceutical industry may step in, it tends to ignore disease types that it considers too risky.

NCATS would allow NIH-funded researchers to create proof-of-concept compounds for neglected-disease targets and biomarkers for drug development, and even to launch early-stage clinical studies. Such 'de-risked' projects would offer in-licensing opportunities for drug firms. The non-profit research community should also benefit (A. J. Stevens *et al. N. Engl. J. Med.* **364,** 535–541; 2011).

Contrary to Miller's view, "government bureaucrats" will not decide which drugs to pursue. NIH programmes slated to move into the NCATS involve researchers at universities, hospitals and institutes, mostly using peer review to allocate funding.

The public is frustrated by the slow translation of basic research into clinical benefits. Venture investors and biotechnology companies are exacerbating this by replacing early-stage research and development with product in-licensing.

John C. Reed* *Sanford-Burnham Medical Research Institute, California, USA.*
jreed@sanfordburnham.org
E. Lucile White* *Southern Research Institute, Alabama, USA.*
**On behalf of 5 co-signatories (see go.nature.com/lcvnnw).*

NIH: FDA key for speed and safety

The US National Institutes of Health (NIH) and the Food and Drug Administration (FDA) should work together to bring products to market (*Nature* **471,** 135; 2011). A telling example reveals how well this can work.

In the early 2000s, the National Institute of Child Health and Human Development conducted clinical trials of 17P, a drug that successfully reduced pre-term birth by more than 30% in women who had experienced it previously (P. J. Meis *et al. N. Engl. J. Med.* **348,** 2379–2385; 2003). The FDA granted the drug orphan status in 2003 to motivate the pharmaceutical company behind 17P to apply for FDA approval, which was granted this year after using the same NIH trial for the approval process.

In the interim, compounding pharmacies sold the drug at about US$400 per course of treatment. Some insurance companies chose not to pay for it because the drug was not FDA approved, which increased their later spend on complications from preterm births that could have been avoided.

With 17P approved and protected from competition for 7 years, the pharmaceutical company raised the drug's price to $30,000 and sent stern letters to pharmacies. In a later announcement of FDA drug-enforcement priorities, 17P was low on the list because no safety issues had been reported regarding pharmacies.

This breakthrough would not have occurred had the NIH not focused on solving the costly and complex issue of preterm birth, or without the FDA having granted orphan status and publicly announcing its enforcement priorities.

James P. Reichmann *American HomePatient, Tennessee, USA.*
jim.reichmann@ahom.com

Slovenian scientists reward mentors

Inspired by *Nature*'s Mentoring Awards (*Nature* **471,** 547; 2011), the Society of Young Researchers of Slovenia is recognizing the country's research mentors with its own award (see go.nature.com/hiddyb).

The society made the first call for nominees in 2009 after gathering information about good mentoring practices and lobbying for support from Slovenia's research institutes and funding bodies. The response exceeded expectations, with 32 nominations for 26 nominees (about 800 mentors were eligible). Nominations remained steady in 2010 (24 nominations and 22 nominees) and more institutions were represented.

Developing and promoting the initiative has paid off. Several prominent research institutions and government bodies, as well as many individual mentors, have expressed their support.

Anže Županič* *Newcastle University, UK.*
anze.zupanic@ncl.ac.uk
**On behalf of 17 co-signatories (see go.nature.com/3a2mwj).*

COMPLEX SYSTEMS

Unzipping Zipf's law

One mathematical model can account for power-law distributions in a variety of systems. Eschewing system-specific assumptions, it utilizes a shared feature of the observed distributions: they all describe the division of items into groups.

LADA ADAMIC

Perhaps the only thing more abundant in both natural and man-made systems than power laws are the models that have been developed to explain them. Writing in the *New Journal of Physics*, Baek *et al.*[1] argue that because such models depend on the specifics of each system, they fail to capture the shared cause of this regularity. The authors instead propose a general model that can be applied to any division of items into groups, and that can, for example, account for Zipf's law of word frequencies in text, the popularity of last names, and city and county populations.

Scientists have been captivated by power laws with reason. Whereas other probability distributions invariably bend on log–log scales, the power law continues as a perfectly straight line over as many orders of magnitude as the system size allows. A power-law distribution, in special cases referred to as Zipf's law or a Pareto distribution, specifies that the probability of observing an item of size k is proportional to $k^{-\alpha}$, with α being typically between 1 and 3. The implications of the distribution are even more striking than its heavy-tailed shape: there are a few megacities but many small towns; a small number of individuals hold a substantial fraction of the total wealth; and there are roughly 2.5 million Smiths in the United States, whereas most last names are uncommon. In fact, a heavy-tailed distribution of sizes tends to hold for a wide range of systems in which items are assigned to bins: species to genera, readers to books, visitors to websites, written words to vocabulary (Fig. 1), and even casualties to wars[2].

Various models have been proposed to explain one or several of the observed power laws. Two main criticisms are commonly aimed at such models. First, many distributions deviate, at least slightly, from a straight line on a log–log scale. Often the deviation is an exponential cut-off in the tail of the distribution and is not captured by the model. Second, models tend to contain system-specific elements that limit their generalizability, and early pursuits of more general models were undertaken by, among others, Herbert Simon, who wrote[3]: "No one supposes that there is any connection between horse-kicks suffered by soldiers in the German army and blood cells on a microscopic slide other than that the same urn scheme provides a satisfactory abstract model for both phenomena."

WORDLE.NET

Figure 1 | Most frequently occurring words in a novel. Words are sized in proportion to their frequency of appearance in *The Mayor of Casterbridge*, a novel by Thomas Hardy that is part of a corpus analysed by Baek and colleagues[1]. The word frequencies follow Zipf's law; the single word 'the' appears 6,775 times in the novel, whereas 4,959 words occur only once, among them 'grouping' and 'mathematical'.

The urn model proposed by Simon is related to other preferential-attachment growth models, also known as cumulative-advantage or 'rich-get-richer' processes. Yule developed[4] one of the oldest such models, proposing that genera grow in proportion to the number of species they contain, by assuming that each species has an equal likelihood of generating a speciation event.

Whereas preferential-attachment models continue to be used to explain power-law distributions observed in various contexts[5,6], some power laws prompt different explanations. For example, Zipf's distribution of word frequencies can result from a principle of least effort[7,8] in the evolution of language, or even from random sequences of letters and spaces[9]. This leaves open the possibility that there is a more general, global explanation of power laws that is independent of system-specific details.

Just such an explanation has been developed by Baek and colleagues[1]. Their random group formation (RGF) model is built on the only common feature among all the systems modelled: that M items are divided among N groups. Entropy is maximized when an item is equally likely to be found at any of M 'addresses' across the groups. Next, one derives a distribution of group sizes that minimizes the amount of information needed to locate an item knowing only the size of the group to which it belongs. This objective, in addition to the constraints of total number of groups and items, and the maximum group size, is sufficient to derive the RGF function, a power-law distribution of group sizes with an exponential cut-off.

There are several remarkable aspects to this finding. The RGF function closely fits observed group size distributions without incorporating any knowledge of system-specific dynamics. In contrast to previous models, which would typically tune their parameters by fitting the empirically observed distribution, the RGF model requires no tuning. The power-law exponent in the RGF function is given directly once one specifies M, N and the maximum observed group size. Furthermore, the exponential cut-off observed in empirical data is an essential component

of the RGF model, rather than a correction introduced to fit the data. The RGF model just as easily fits word-frequency distributions representing entire books as it fits random subsamples of the same texts, something that alternative models generally cannot do. Finally, the approach is flexible enough to incorporate system-specific constraints, as needed.

The work of Baek and colleagues[1] may be the first to provide a truly general explanation of the prevalence of power-law distributions in frequency counts. But it is not yet ready to replace other models entirely. For many, if not all, systems the intuition behind the assumption that one wishes to minimize the information cost of locating an item needs to be further developed. By contrast, growth models usually integrate intuition about a system's evolution. Furthermore, the power-law exponents produced by the RGF model in some cases differ from those estimated previously using maximum-likelihood fits to data[2]. Nevertheless, by deriving power-law distributions from very general system-independent principles, Baek *et al.* have raised the bar for other models. A model purporting to explain a power-law distribution should be as general as Baek and colleagues' model, or it should be able to reproduce additional features of the system it models, beyond the familiar straight line on a log–log plot. ■

Lada Adamic *is in the School of Information and the Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan 48109-1107, USA.*
e-mail: ladamic@umich.edu

1. Baek, S. K., Bernhardsson, S. & Minnhagen, P. *New J. Phys.* **13,** 043004 (2011).
2. Newman, M. E. J. *Contemp. Phys.* **46,** 323–351 (2005).
3. Simon, H. A. *Biometrika* **42,** 425–440 (1955).
4. Yule, G. U. *Phil. Trans. R. Soc. Lond. B* **213,** 21–87 (1925).
5. Barabási, A.-L. & Albert, R. *Science* **286,** 509–512 (1999).
6. Huberman, B. A. & Adamic, L. A. *Nature* **401,** 131 (1999).
7. Zipf, G. K. *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology* 1st edn (Hafner, 1949).
8. Ferrer i Cancho, R. & Solé, R. V. *Proc. Natl Acad. Sci. USA* **100,** 788–791 (2003).
9. Miller, G. A. *Am. J. Psychol.* **70,** 311–314 (1957).

STEM CELLS

iPS cells under attack

Induced pluripotent stem cells offer promise for patient-specific regenerative therapy. But a study now cautions that, even when immunologically matched, these cells can be rejected after transplantation. SEE LETTER P.212

EFFIE APOSTOLOU & KONRAD HOCHEDLINGER

In 2006, Takahashi and Yamanaka[1] made a groundbreaking discovery. When they introduced four specific genes associated with embryonic development into adult mouse cells, the cells were reprogrammed to resemble embryonic stem cells (ES cells). They named these cells induced pluripotent stem cells (iPS cells). This approach does not require the destruction of embryos, and so assuaged the ethical concerns surrounding research on ES cells[2]. What's more, researchers subsequently noted that the use of 'custom-made' adult cells derived from human iPS cells might ultimately allow the treatment of patients with debilitating degenerative disorders. Given that such cells' DNA is identical to that of the patient, it has been assumed — although never rigorously tested — that they wouldn't be attacked by the immune system[3]. On page 212 of this issue, however, Zhao *et al.*[4] show, in a mouse transplantation model, that some iPS cells are immunogenic, raising concerns about their therapeutic use.

To examine the immunogenicity of mouse iPS cells, the authors used a simple teratoma-formation assay. Briefly, they injected iPS cells into mice that were either immune-compromised or genetically matched with the donor cells. This normally results in the formation of benign tumours called teratomas, which consist of many types of differentiated cells. Zhao *et al.* validated their approach by showing that a line of genetically matched (autologous) ES cells gives rise to teratomas, whereas a line of unmatched ES cells is rejected by the immune system of the recipient animals before it can produce teratomas (Fig. 1a).

Surprisingly, transplantation of autologous iPS cells derived from fetal fibroblasts into matched mice also resulted in the rejection of teratomas (Fig. 1b). To rule out the possibility that the viral vectors used to introduce the reprogramming genes (which integrate into the host-cell genome) were responsible for the immune rejection, the authors used a different method — the episomal approach — to generate iPS cells. This produced similar results, albeit with a weakened immune response. Together, these data indicate that, in this assay, matched iPS cells are more immunogenic than matched ES cells.

Zhao *et al.*[4] further identified the antigens that probably caused immune rejection of the iPS cells. By analysing the gene-expression profiles of iPS-cell-derived teratomas, they discovered a group of just nine genes that were expressed at abnormally high levels. Indeed, inducing expression of three of these genes (*Hormad1*, *Zg16* and *Cyp3a11*) in the non-immunogenic ES cells significantly impaired these cells' ability to form teratomas on transplantation into genetically matched mice. Whether activation of the same genes is also necessary for immunogenicity of iPS cells remains to be tested.

Teratoma regression may be due to a

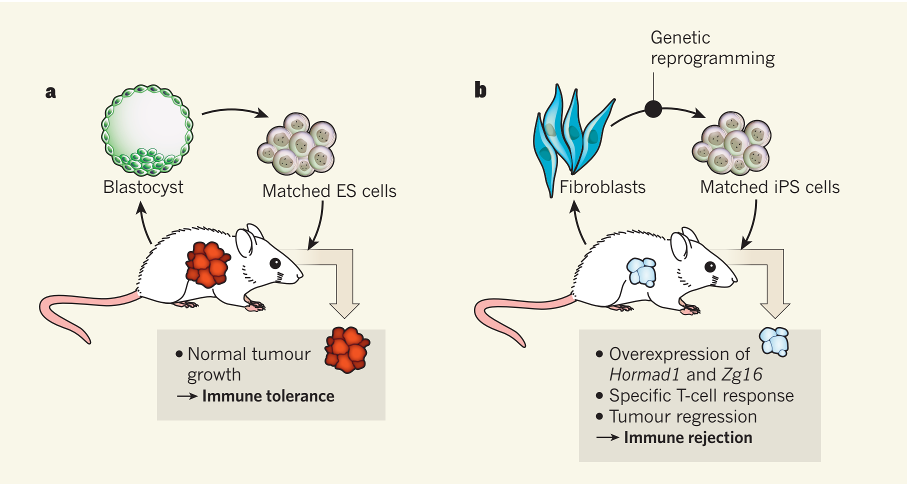


Figure 1 | Immunogenicity of induced pluripotent stem cells. a, Zhao *et al.*[4] find that embryonic stem cells (ES cells) derived from blastocyst embryos from a given genetic background grow into teratomas on transplantation into mice of the same genetic background. The immune system therefore 'tolerates' autologous ES cells. **b,** By contrast, autologous induced pluripotent stem cells (iPS cells), reprogrammed from fetal fibroblasts by viral or non-viral genetic approaches, elicit an unexpected immune reaction in genetically identical mice, resulting in their rejection.

specific T-cell-driven immune response that is normally seen during organ rejection. To test this possibility, Zhao and co-workers injected iPS cells into mice lacking subsets of T cells — either $CD4^+$ T-helper cells or $CD8^+$ cytotoxic T cells. Under these conditions, teratomas could grow. The authors also show that the immunogenicity of iPS-cell-derived teratomas results from a specific T-cell response to products of *Hormad1* and *Zg16*.

The observation that even autologous mouse iPS cells are rejected by a specific immune response in matched hosts is quite unexpected. But although these findings serve as a cautionary note for the ultimate therapeutic use of iPS cells, many questions remain.

For example, it is unclear why and how the nine genes become activated in iPS cells but not in ES cells. It has recently been suggested[3] that the reprogramming process introduces epigenetic and genetic abnormalities into iPS cells, which is one conceivable way in which gene-expression patterns could be perturbed. In addition, an influence of the cell type of origin on the immunogenic properties of resultant iPS cells must be explored, given that iPS cells from different somatic cells can exhibit distinct molecular and functional properties[5,6]. Most importantly, it should be determined whether the different types of differentiated cells derived from iPS cells will also elicit an immune response after transplantation.

Indeed, immature iPS cells as used in this study[4] would never be used for transplantation in a clinical setting, and it is unclear whether tumour antigens would continue to be expressed after differentiation into therapeutically relevant cell types. Notably, previous work has shown[7] that haematopoietic progenitors derived from iPS cells and matured *in vitro* can successfully engraft in genetically matched mice without obvious rejection, although the recipient animals had to be irradiated first.

Another key question is to what extent the reprogramming method might influence the immunogenicity of resultant iPS cells. The authors' observation that iPS cells generated using the episomal approach are less immunogenic than those generated using viral vectors is promising, because it suggests that alternative methods may entirely prevent immunogenicity.

Lastly, whether human iPS cells trigger immunogenicity must be investigated. A way to do this is to test whether teratomas derived from human iPS cells are rejected in mice carrying a genetically matched, humanized immune system.

Regardless of the answers to the outstanding questions, this and other recent studies reach one common conclusion: researchers must learn more about the mechanisms underlying cellular reprogramming and the inherent similarities and differences between ES cells and iPS cells. Only on careful examination of these issues can we know whether such differences pose an impediment to the potential therapeutic use of iPS cells, and how to address them. In any case, these findings should not affect the utility of iPS-cell technology for studying diseases and discovering drugs *in vitro*[3]. ■

Effie Apostolou *and* **Konrad Hochedlinger** *are at the Howard Hughes Medical Institute, Center for Regenerative Medicine and Cancer Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. They are also in the Department of Stem Cell and Regenerative Biology, and at the Harvard Stem Cell Institute, Harvard University.*
e-mail: khochedlinger@helix.mgh.harvard.edu

1. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
2. Stadtfeld, M. & Hochedlinger, K. *Genes Dev.* **24,** 2239–2263 (2010).
3. Wu, S. M. & Hochedlinger, K. *Nature Cell Biol.* **13,** 497–505 (2011).
4. Zhao, T., Zhang, Z.-N., Rong, Z. & Xu, Y. *Nature* **474,** 212–215 (2011).
5. Kim, K. *et al. Nature* **467,** 285–290 (2010).
6. Polo, J. M. *et al. Nature Biotechnol.* **28,** 848–855 (2010).
7. Hanna, J. *et al. Science* **318,** 1920–1923 (2007).

K.H. declares competing financial interests. See go.nature.com/w6v7kl for details.

EPIDEMIOLOGY

How common is autism?

Autism spectrum disorders vary greatly in severity. By including children in regular education who received no special help, an epidemiological study has found these disorders to be up to three times more prevalent than thought.

CATHERINE LORD

Autism spectrum disorders (ASDs) encompass autism, Asperger's syndrome and other related conditions. Core features include difficulties in basic social and communicative behaviours such as eye contact, intonation and facial expressions; and repetitive behaviours and restricted interests, from finger mannerisms to preoccupation with heating vents. The prevalence of ASDs has been considered to be just over 1% (ref. 1). However, in a provocative, carefully executed study published in *The American Journal of Psychiatry*, Kim *et al.*[2] present evidence for surprisingly high rates of these disorders — 3.74% in males and 1.47% in females — in school-age children in a South Korean community. Remarkably, more than two-thirds of the children diagnosed with an ASD in this study were receiving no special help and had no previous psychiatric or developmental diagnoses.

Although ASDs are defined by the shared difficulties mentioned above, anyone who has met more than one person with an ASD is struck by the differences between these individuals. With eye contact, for example, there may be active avoidance, staring for too long or a lack of the usual timing of gaze. Variations in autism severity, cognitive skills and language levels are apparent even between family members — including identical twins — who have autism[3]. In the search for explanations for this variation, epidemiological studies are essential and can also identify factors that might increase the risk of developing an ASD — especially given that, despite several hypotheses about neurobiological pathways underlying these disorders[4], no diagnostic biomarkers have been identified. Knowing the prevalence of autism in different geographical regions and cultures is important for comparing variation and sources of risk .

However, epidemiological research in autism faces many challenges. To recruit subjects for a study of this kind, researchers depend either on medical or educational records, or on behavioural descriptions — usually measured through parent or teacher questionnaires. To minimize the stigma associated with ASDs, Kim *et al.*[2] described their project to parents and teachers as being about "difficulties in social relationships".

Most children with poor social relationships do not have an ASD. Consequently, ASD screening questionnaires typically misidentify substantial numbers of children who have other difficulties, but not an ASD[5]. The authors addressed this concern by providing comprehensive diagnostic assessments for children who screened positive, using Korean translations of internationally accepted diagnostic interviews and observations. Moreover, they paid substantial attention to cross-cultural issues in translations, and to documenting consistent, standard behavioural coding across examiners. What's more, North American team members used their judgement to confirm the South Korean diagnoses. However, the translated scales were not pre-tested by independent experts in South

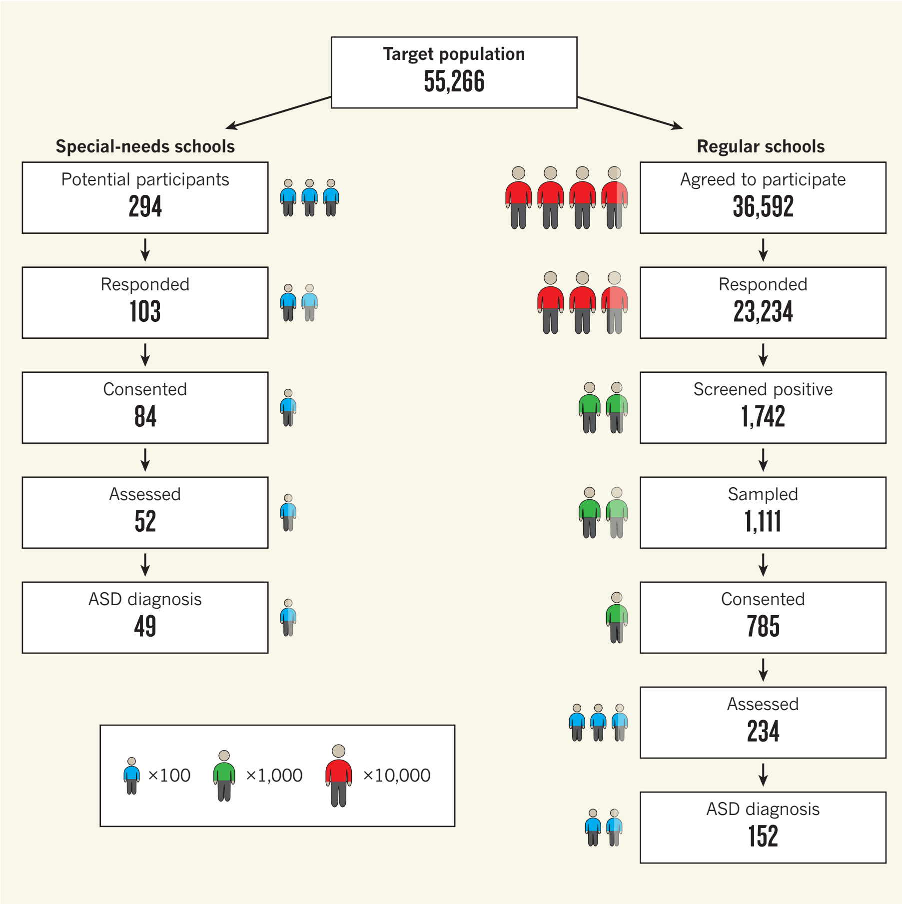


Figure 1 | Kim and colleagues' study[2]. Although the authors set out to investigate a large target population of 7–12-year-old children from two groups in a South Korean community, the number of children they eventually assessed for autism spectrum disorders (ASDs) gradually decreased, as is typical of such studies.

Korea, to show that they could differentiate children diagnosed with ASDs from those without — the standard way in which the methods are evaluated in other cultures.

The study is remarkable in its attempt to identify ASDs in children in mainstream schools, as well as in children receiving special care. The overall response rate was about 60% (Fig. 1). Although 7% of the mainstream schoolchildren screened positive, only 13.4% of these could be further evaluated, primarily because families did not consent to, or did not attend, subsequent assessments. In the end, Kim and colleagues assessed a total of 286 children, and used a range of statistical measures to reach estimates for a population of more than 55,000. In such research, prevalence rates depend on data from the screening questionnaires, which are necessary to make inferences about children not seen. The priority is that no case should be missed. The negative consequences — both clinical and political — of low rates are often assumed to be much greater than those of overestimating.

Yet a larger question arises. Should individuals who are without impairment or disability — that is, without any suffering, limitations or restrictions in daily functioning — be diagnosed with an ASD? In many areas of medicine, diagnosis with a disease (for instance, diabetes) can be clearly separated from the associated disability (from no effect to significant limitation in activity)[6]. By contrast, for a variety of reasons — including avoiding the unfortunate stigma of psychological difficulties — standard-setting agencies[7] have typically required that definitions of ASDs include an impairment or, in the case of young children, a high risk of impairment before a diagnosis is considered.

To identify the presence of impairment, many studies include a measure of adaptive functioning that indicates children's ability to carry out everyday tasks; this is usually markedly impaired in those with ASDs, even in children of high intelligence[8]. In the South Korean study, however, it was not clear whether the children identified as having an ASD needed help but were not receiving it, or whether the mainstream schoolchildren estimated to have considerable ASD symptoms were otherwise without impairments. (The latter would possibly support proposals by self-advocacy groups[9] and others[10] that ASDs are conditions not disorders, but contradicts norms for most standard

NATURE

50 Years Ago

'[A Symposium on] Solar Variations, Climatic Change and Related Geophysical Problems' — It became abundantly clear how large a number of investigators are patiently accumulating evidence of the amplitude, character, effects and especially the dating of climatic fluctuations all over the world. Speculations regarding the causes abound; supporters of each of the popular theories — solar variation, atmospheric turbidity, carbon dioxide, ozone, variations in the Earth's orbital elements — find their several gods alternately set up and cast down … It now seems almost certain that no one simple panacea is in any way adequate to explain the intriguing patterns of climatic fluctuation to which the evidence points.

From *Nature* 10 June 1961

100 Years Ago

In the meteorological charts of the North Atlantic and Indian Oceans for the month of June … there is an interesting article on the phenomenon of St. Elmo's Fire or corposants (*corpus sanctum*), the harmless luminous electricity of low intensity seen sometimes at night on ships' masts, &c., during unsettled weather. Many examples of authenticated experiences in olden and modern times are quoted, *e.g.* one by Columbus in October, 1495, during a severe storm. It was then assumed that the light emanated from the saint's body, and was a sure sign that the gale was at its maximum … The phenomenon is not unfrequent on land; it was quoted by Caesar and others. On the summit of Ben Nevis the observatory was at times ablaze with it; the observers were not in any way inconvenienced, except by a slight tickling sensation in head and hands.

From *Nature* 8 June 1911

therapeutic or educational classifications.)

The diagnostic procedures Kim *et al.* used have consistently been shown to distinguish between children with ASDs and those with other disorders[11]. However, systematic comparisons have not been made between children diagnosed with an ASD by a clinician and children thought to have social–communication deficits on the basis of parent or teacher reports, but whom experts consider not to have an ASD.

It is possible that diagnostic instruments are less effective in making this latter distinction than we might expect. So in breaking new ground, Kim and co-workers' study[2] cries out for replication. It provides a challenge to expand research across cultures, to think critically about biases that could affect inferences in the study of total population samples, and to consider the possibility of undiagnosed ASDs in children in ordinary classrooms, with or without impairments. ■

Catherine Lord *is in the University of Michigan Autism and Communication Disorders Center, Ann Arbor, Michigan 48109, USA.*
e-mail: celord@umich.edu

1. Fombonne, E. *Pediatr. Res.* **65,** 591–598 (2009).
2. Kim, Y. S. *et al. Am. J. Psychiatry* doi:10.1176/appi.ajp.2011.10101532 (2011).
3. Le Couteur, A. *et al. J. Child Psychol. Psychiatry* **37,** 785–801 (1996).
4. Moldin, S. O. & Rubenstein, J. L. R. (eds) *Understanding Autism* (Taylor & Francis, 2006).
5. Ehlers, S., Gillberg, C. & Wing, L. *J. Autism Dev. Disord.* **29,** 129–141 (1999).
6. www.who.int/icidh
7. www.DSM5.org
8. Kanne, S. M. *et al. J. Autism Dev. Disord.* doi:10.1007/s10803-010-1126-4 (2010).
9. www.aspiesforfreedom.com
10. Baron-Cohen, S. *et al. Br. J. Psychiatry* **194,** 500–509 (2009).
11. Risi, S. *et al. J. Am. Acad. Child Adolesc. Psychiatry* **45,** 1094–1103 (2006).

The author declares competing financial interests. See go.nature.com/wdjx6a for details.

PHYSICAL CHEMISTRY

Water's wafer-thin surface

A combination of vibrational spectroscopy and molecular calculations reveals that only the surface layer of water at the interface with air has a distinctly different structure from the bulk liquid. SEE LETTER P.192

PAVEL JUNGWIRTH

The next time you drink a glass of water, take a moment to consider the surface of the liquid. It might surprise you to learn that the molecular structure of this seemingly unremarkable air–water interface has long been the subject of controversial theories. On page 192 of this issue, Stiopkin *et al.*[1] weigh into the debate by reporting the results of spectroscopic and theoretical studies of this interface. Their findings strongly contradict the recurring idea that external interfaces with water, or compounds dissolved in it, can generate long-range molecular order in the liquid.

There have been two popular, mutually interconnected misconceptions about liquid water. According to the first, water can 'remember', for extended periods of time, patterns imposed on it by the environment or by solute molecules[2]. This idea, among others, lent credence to a central tenet of homeopathy: that solutions of compounds could maintain biological activity even when repeatedly diluted to such an extent that effectively no molecules of those compounds remained in solution. But the 'memory of water' effect is not reproducible[3] and the idea is now scientifically unacceptable — although this doesn't yet seem to have affected the commercial success of homeopathy.

The second misconception is that long-range order across hundreds or even thousands of molecular layers can exist in liquid water. An early incarnation of this idea was the report[4] of polywater — a syrupy liquid apparently produced when water was repeatedly forced through capillary tubes, and which was thought to be polymerized water. As with the 'memory of water', the existence of polywater has been debunked[5], but similar ideas have proven to be more durable in the scientific literature. Indeed, there has been a renaissance of the notion that long-range order can be imposed on water by an interface of the liquid with air or with another immiscible liquid, or by an interface with a solid such as the containing vessel[6].

Moreover, the dust has been blown off the picture of bulk water as a mixture of regions of two distinct liquid forms, a concept first proposed[7] heuristically by Wilhelm Röntgen more than a century ago. In the most recent variant of this idea[8], robust X-ray scattering data for liquid water have been fitted into a microscopic version of Röntgen's elderly model. And, in keeping with the theory that liquid water is structured by surfaces, it has been proposed that long-range water patterning occurs around proteins and other biomolecules. This in turn has led to the enigmatic concept of 'cellular water'[9] — the idea that all water molecules in cells are ordered, and that this order is crucial for biological function.

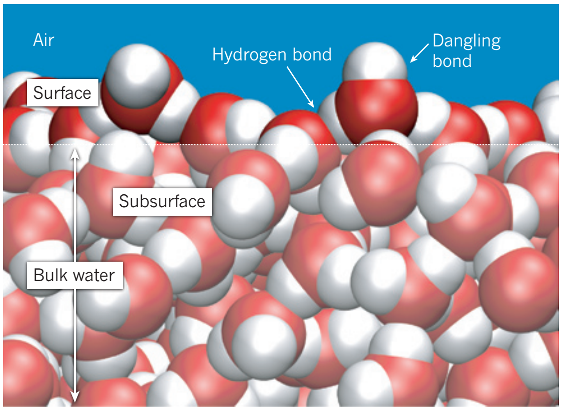


Figure 1 | Structure of the interface between liquid water and air. This image, taken from a classical molecular-dynamics simulation, reveals the features of liquid water at an air–water interface. Most water molecules form hydrogen bonds to each other, but some at the surface do not (these are known as dangling bonds). Stiopkin *et al.*[1] report evidence that the surface has a depth of only one layer of molecules, contradicting theories that interfaces with water impose long-range order deep into the bulk liquid. The surface layer of water molecules thus sits atop a subsurface layer, the properties of which approach those of bulk water. Red atoms are oxygen, white ones are hydrogen.

Stiopkin *et al.*[1] have therefore performed a valuable service by looking in great detail and with little prejudice at the structure of an archetypal aqueous interface, that between water and air. They posed a seemingly simple question: how thick is this interface? In other words, how deep into bulk water does the patterning imposed by the surface propagate? In fact, for liquid water, with its highly dynamical and volatile surface, this is a difficult question to answer.

The authors succeeded by applying a combination of two state-of-the-art techniques, which they also helped to develop. The first was experimental: they used a sensitive version of surface-selective vibrational spectroscopy to study oxygen–hydrogen bonds in water that stick out into the air, and that are therefore not hydrogen-bonded to another water molecule. (Actually, for technical reasons, they studied oxygen–deuterium bonds of water molecules isotopically labelled once or twice with

deuterium.) Such free or dangling bonds (Fig. 1) exist only at the surface and can thus be used as sensitive 'antennas' for detecting interactions of surface water molecules with molecules in the underlying subsurface water layer.

Stiopkin and colleagues' second approach was to perform sophisticated and accurate calculations to model the vibrations of water molecules in the liquid. This allowed them to compare the strength of hydrogen bonds between water molecules in the interfacial layer and in the bulk. They also used an old but neat trick, known as isotopic dilution, to disentangle the contributions of intramolecular and intermolecular couplings of individual water molecules to observed spectral shifts. This involved analysing a set of samples prepared by mixing together varying fractions of ordinary water (H_2O) and doubly deuterated water (D_2O).

By now, the reader may be waiting impatiently for the great news concerning amazing structural patterns discovered at the surface of water. If so, the reader will be disappointed — Stiopkin *et al.* found that the surface does not have a long-range impact on water and that the strength of interactions of surface water molecules with those in the subsurface layer is comparable to that between water molecules in the bulk. This means that only the surface layer is distinctly different in structure from the rest of the liquid, and that the water surface has the thickness of just one layer of water molecules, which is about 0.3 nanometres.

Other studies have shown that the surface thickness of water may be extended by more complex interfaces (such as charged ones[10]) or by solutes[11], but not to more than a small number of water layers. When taken together with those previous studies, the main message of Stiopkin and colleagues' findings[1] is that the old models of long-range patterning in liquid water and their newer incarnations are not substantiated. The fact that these models are not always perceived merely as historical curiosities warrants broad attention to the present work. The authors' findings may seem to represent a 'negative' result. But by offering a sober view of the fascinating and important — but by no means magical — properties of liquid water, they should have a most welcome, positive effect. ■

Pavel Jungwirth *is at the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 16610 Prague 6, Czech Republic.*
e-mail: pavel.jungwirth@uochb.cas.cz

1. Stiopkin, I. V. *et al. Nature* **474,** 192–195 (2011).
2. Davenas, E. *et al. Nature* **333,** 816–818 (1988).
3. Ball, P. *Nature* doi:10.1038/news070806-6 (2007).
4. Deryagin, B. V. & Fedyakin, N. N. *Dokl. Akad. Nauk. SSSR* **147,** 403–405 (1962).
5. Rousseau, D. L. & Porto, S. P. S. *Science* **167,** 1715–1719 (1970).
6. Zheng, J., Chin, W., Khijniak, E., Khijniak, E. Jr & Pollack, G. H. *Adv. Colloid Interface Sci.* **127,** 19–27 (2006).
7. Röntgen, W. K. *Ann. Phys. Chem. N.F.* **45,** 91–97 (1892).
8. Huang, C. *et al. Proc. Natl Acad. Sci. USA* **106,** 15214–15218 (2009).
9. Pollack, G. H. *Cells, Gels and the Engines of Life* (Ebner, 2001).
10. Jena, K. C., Covert, P. A. & Hore, D. K. *J. Phys. Chem. Lett.* **2,** 1056–1061 (2011).
11. Gopalakrishnan, S., Jungwirth, P., Tobias, D. J. & Allen, H. C. *J. Phys. Chem. B* **109,** 8861–8872 (2005).

NEUROSCIENCE

Flies race to a safe place

Is the fruitfly one of the animals that use visual cues to guide navigation and spatial learning? And if so, how does the simple fly brain support spatial memory formation? Some answers are now emerging. **SEE LETTER P.204**

TROY ZARS

Visual cues can guide our ability to navigate and help us to form spatial memories. The importance of such cues is most obvious when the task is difficult, say when retracing a route to a distant relative's house without the help of the Global Positioning System. The use of visual cues for spatial memory is not unique to humans. Impressively, even insects such as ants and honeybees use them to find food sources, and the home nest following what are sometimes quite long foraging trips[1,2]. On page 204 of this issue, Ofstad *et al.*[3] investigate whether and how the fruitfly *Drosophila melanogaster* — which has no obvious home, and therefore perhaps no driving force to develop such a system — can use visual landmarks for navigation and the formation of spatial memory.

Spatial memory in *Drosophila* has been examined in simple as well as more complex environments. Perhaps the simplest paradigm is one that allows individual flies to walk back and forth inside a long, narrow, dark chamber (a heat box)[4]. Spatial memory is induced by automatically pairing a high temperature with the movement of a fly to one part of the chamber. The fly can use this place–temperature pairing to form a memory that lasts hours, presumably using spatial information that it has somehow derived.

To test for an orientation memory, a more complicated set-up is needed[5]. In this case, if a fly is allowed to roam on an open disk placed between two conspicuous vertical stripes, the insect will walk back and forth between the two landmarks for hours. If, however, a distracting landmark is presented transiently for several seconds, the fly will orient towards this new stripe. When all landmarks are then removed, the insect will reorient towards the original — but now absent — visual target, even though there are no obvious rewards or punishments as an incentive for it to do so.

Ofstad *et al.*[3] extend these studies by asking whether flies can use conspicuous visual landmarks to form more-stable spatial memories in a complex open-field environment. For this, they used a novel set-up inspired by tests of spatial memory that were developed for rodents and crickets[6,7]. The set-up, which in some ways is similar to one developed independently by a different group[8], uses a mosaic platform of multiple tiles (Fig. 1). The authors could heat the tiles to an aversive 36 °C, leaving only a single tile at the preferred temperature of 25 °C. They also supplied visual cues — stripes of different orientations — by surrounding the platform with arrays of light-emitting diodes. During conditioning, a computer fixed the position of the cool tile with respect to the visual landmarks such that although the position of the cool tile changed, the orientation of the stripes with respect to it was constant.

After the flies had undergone multiple training trials, Ofstad *et al.* performed a spatial memory test by shifting the visual panorama clockwise or anticlockwise by 90° and raising the temperature of all tiles to the aversive temperature. They tracked the position of individual flies by camera throughout the experiment to correlate spatial preference with both the location of the cool tile during training and the expected position of the cool tile, even though it was not there after the 90° shift.

Impressively, flies searched for about a minute in the quadrant where they expected the cool tile to be. To test for the duration of this spatial memory, Ofstad and colleagues took the flies out of the arena and then returned them at various times later. The memory lasted for about 2 hours. In control experiments, flies could still find the cool tile in the dark or when visual-cue conditions were uncoupled from temperature over several subsequent trials. But, as expected, there was no preference for the target quadrant when the temperature was raised throughout the platform in a memory test.

What is especially impressive about the spatial memory examined here is that it presumably involves a more complex process than any other type of memory tested in *Drosophila*. In other tests, flies essentially choose between one of two options (for instance, walk or fly to the

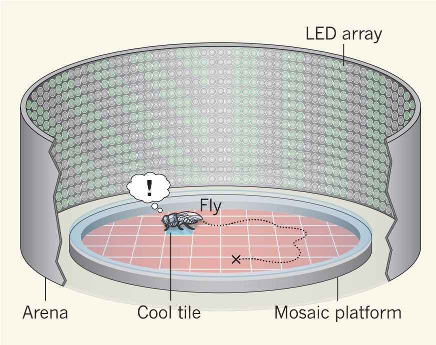


Figure 1 | Model for studying visually guided spatial memory[3]. Individual flies are free to roam on a platform made of heated tiles. Visual patterns on the surrounding walls — generated by arrays of light-emitting diodes (LEDs) — provide orientation cues to guide navigation to a single cool tile. Flies remember the position of this safe tile after the visual patterns have been rotated by 90° and the platform is uniformly heated.

left or right)[9]. In Ofstad and colleagues' test, the insects can move in any direction on the disk to reach the cool tile.

Ofstad *et al.*[3] also investigated what parts of the fly brain are required for spatial memories. Much effort has focused on the role of the mushroom bodies in memory formation; these are a pair of structures in the insect brain that are involved in some forms of learning and memory. For example, olfactory memories — which depend on the association of one of two odorants with appetitive or aversive cues and which are analysed by an orientation test between the two odorants — are formed in the mushroom bodies[10]. In cockroaches, cutting the mushroom bodies with foil blades hinders the formation of a spatial memory that also uses visual landmarks[6]. However, studies in *Drosophila* have shown[9] that the simple spatial memory tested using the heat-box paradigm does not require these structures.

Using several techniques, Ofstad and co-workers[3] show that the mushroom bodies are not required for visually guided spatial memory either. Instead, they find that altering the ellipsoid body — part of the four-component central complex that has been associated with premotor functions (involving voluntary movement) and some other types of memories — strongly reduced the flies' ability to form spatial memories. These results are consistent with the idea that there is no single centre in the fly brain for memory formation, and that specialized neural systems are crucial for different forms of memory.

In addition to the identification of parts of the nervous system that are essential for spatial memory, the authors' design of a rapid, visually guided spatial memory test in *Drosophila* raises the exciting prospect of genetic analyses of this form of memory. If results from previous comparative genetic studies[9] are anything to go by, one should expect that only some genes will have common roles in memory formation across tasks, with others having functions of greater specificity. Identification of the genes and neural circuits that are crucial for visually guided spatial memory is an exciting prospect indeed. ■

Troy Zars *is in the Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211, USA.*
e-mail: zarst@missouri.edu

1. Menzel, R. *et al. Curr. Biol.* **21,** 645–650 (2011).
2. Collett, M. *Proc. Natl Acad. Sci. USA* **107,** 11638–11643 (2010).
3. Ofstad, T. A., Zuker, C. S. & Reiser, M. B. *Nature* **474,** 204–207 (2011).
4. Zars, T. *J. Neurogenet.* **23,** 104–110 (2009).
5. Neuser, K. *et al. Nature* **453,** 1244–1247 (2008).
6. Mizunami, M., Weibrecht, J. M. & Strausfeld, N. J. *J. Comp. Neurol.* **402,** 520–537 (1998).
7. Morris, R. *J. Neurosci. Meth.* **11,** 47–60 (1984).
8. Foucaud, J., Burns, J. G. & Mery, F. *PLoS ONE* **5,** e15231 (2010).
9. Zars, T. *Learn. Mem.* **17,** 246–251 (2010).
10. van Swinderen, B. *Curr. Biol.* **19,** R855–R857 (2009).

QUANTUM PHYSICS

How to catch a wave

The wavefunction is a central mathematical entity in quantum physics. It is used to completely describe the state of a system. A means of probing it directly is now on hand. SEE LETTER P.188

ONUR HOSTEN

Quantum mechanics is the framework for describing the physics of the microscopic world. Central to this description is the wavefunction, which contains all the information about the relevant physical system. To date, experimental determination of wavefunctions has been accomplished only through inferences based on indirect measurements. But that has now changed. On page 188 of this issue, Lundeen *et al.*[1] present a method to measure the wavefunction directly.

The complex-valued wavefunction associated with a quantum system is not itself considered to be a physical element of quantum theory. Nevertheless, its absolute square, for instance, represents a probability distribution associated with particular outcomes of an experiment; for example, the outcome of finding a particle at a certain location. In this context, a question naturally arises: despite its abstract existence, can an unknown wavefunction of a system be determined experimentally? With just a single copy of the system in hand, this turns out to be impossible — even in principle — owing to the random disturbance that the measurement process imposes on the system[2]. However, with an ensemble of identically prepared systems, it is possible to determine the wavefunction.

By making a set of measurements of each of several different physical properties on the ensemble of identically prepared systems, and using the obtained probability distributions associated with these properties, the sought-after wavefunction can be constructed algorithmically. This indirect way of characterizing the wavefunction is known as quantum-state tomography[3,4], and it has been a quintessential tool in the field of quantum-information science. By contrast, Lundeen and colleagues' method[1] directly probes the real and imaginary parts of the wavefunction of the ensemble, as they demonstrate with measurements carried out on the transverse spatial wavefunction of single photons.

The key to their technique is the concept of weak quantum measurements. In a generic quantum measurement, the system to be measured is first coupled to another system, the meter, and information about a property of the system, the observable, is acquired from the meter. The system–meter coupling moves the pointer of the meter by different amounts for different states of the observable, and the initial location of the pointer contains some quantum uncertainty. The measurement is said to be a strong one, after the system–meter interaction is over, if the pointer states corresponding to different states of the observable move away from one another by more than the initial uncertainty on the pointer. A weak measurement is simply the case in which the relevant pointer states still overlap to a large extent, yielding little information about the system and disturbing it insignificantly in a single measurement.

Weak measurements take on a new life when combined with post-selection — that is, when they are conditioned on the outcome of a following strong measurement. Prior to post-selection, the centre of the pointer shows the average value of the measured observable. Following post-selection, owing to an interference effect, the pointer shifts to a new value called the weak value of the observable[5]. Note that ascertaining a weak value requires many repetitions of the same measurement on identically prepared systems, so that the pointer's centre can be identified. In the past two decades, weak values have been used extensively to analyse certain quantum paradoxes, for example Hardy's paradox[6], and most recently

they have led to quite useful techniques for measuring small signals[7,8].

At the heart of Lundeen and colleagues' method[1] lies the observation that a weak measurement of a particle's position followed by a strong measurement of its momentum should yield the particle's spatial wavefunction as the weak value, provided that the measured momentum is zero. A photon's position along an axis (x-axis) transverse to a chosen central propagation axis (z-axis) is no exception to this argument. In their experiment, Lundeen *et al.* obtain single photons by means of a process known as spontaneous parametric down-conversion, and, with various optical elements, shape the to-be-determined transverse spatial wavefunction ($\Psi(x)$, where x is the spatial position) of the photons. The meter they use for the weak measurements is the polarization of the very same photons, which serves as a qubit (two-level) meter[9]. A narrow piece cut from an optical element called a waveplate placed at position $x = x_0$ (say at $z = 0$) implements the weak-measurement coupling, rotating very slightly the polarization of the photons if they propagate through this location.

Post-selection of photons with zero momentum is accomplished by first sending the photons through a lens, and then, at the focal plane, blocking all the photons but the ones at position $x = 0$ with a narrow slit. After this stage, an analysis of the polarization of the remaining photons yields $\Psi(x_0)$. In particular, by convention, the average rotation of the polarization is proportional to the real part of $\Psi(x_0)$, and the average change in the ellipticity of the polarization is proportional to the imaginary part of $\Psi(x_0)$. This procedure is repeated for different positions x of the waveplate to map out the complete wavefunction $\Psi(x)$ at $z = 0$. Lundeen and colleagues show that the described procedure works reliably.

The authors' finding — as I phrase it colloquially, that a wavefunction meter can be built to probe wavefunctions, almost like a voltmeter (or oscilloscope) is used to measure voltages — is conceptually rather surprising. Beyond philosophical issues, the results represent a practical finding: their method can be used as a tool in a wide range of fields, from optical to atomic to solid-state physics, all of which are touched on by quantum-information science. But whether the current method can be a viable alternative to quantum-state tomography is yet to be explored. This will require testing if the system–meter coupling can be practically realized in various physical systems and circumstances. It would be interesting to see this work extended to wavefunctions of multi-particle entangled quantum states. ■

Onur Hosten *is in the Physics Department, Stanford University, Stanford, California 94305, USA.*
e-mail: hosten@stanford.edu

1. Lundeen, J. S., Sutherland, B., Patel, A., Stewart, C. & Bamber, C. *Nature* **474,** 188–191 (2011).
2. Alter, O. & Yamamoto, Y. *Quantum Measurement of a Single System* (Wiley, 2001).
3. Raymer, M. G. *Contemp. Phys.* **38,** 343–355 (1997).
4. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
5. Aharonov, Y., Albert, D. Z. & Vaidman, L. *Phys. Rev. Lett.* **60,** 1351–1354 (1988).
6. Lundeen, J. S. & Steinberg, A. M. *Phys. Rev. Lett.* **102,** 020404 (2009).
7. Hosten, O. & Kwiat, P. *Science* **319,** 787–790 (2008).
8. Dixon, P. B., Starling, D. J., Jordan, A. N. & Howell, J. C. *Phys. Rev. Lett.* **102,** 173601 (2009).
9. Wu, S. & Molmer, K. *Phys. Lett. A* **374,** 34–39 (2009).

DNA REPAIR

Cyclin D1 multitasks

Cyclin D1 is one of the drivers of the cell cycle, and its deregulation may promote the development of tumours. Surprisingly, this protein also mediates the repair of damaged DNA, a mechanism that commonly prevents cancer. SEE LETTER P.230

JIRI BARTEK & JIRI LUKAS

The maintenance of genome integrity is a fundamental biological process. A complex network of proteins detects damaged DNA, signals this detection and repairs the damage, to prevent life-threatening diseases such as cancer[1]. This machinery is particularly crucial in cells going through the cell-division cycle, a proliferative process that can go awry in various cancers[1,2]. But the orchestration of the DNA-damage response and the cell cycle is far from understood, despite the key roles of the two processes in cell physiology and pathology. An exciting report by Jirawatnotai and colleagues in this issue[3] (page 230) sheds new light on the matter, identifying an unexpected function for the cell-cycle protein cyclin D1 in DNA repair.

Cyclin proteins drive the cell cycle in partnership with a family of catalytic proteins called cyclin-dependent kinases (CDKs). Various cyclin–CDK complexes fuel the highly regulated progression through the G1, S, G2 and M phases of the cycle by phosphorylating — thereby activating or inactivating — a

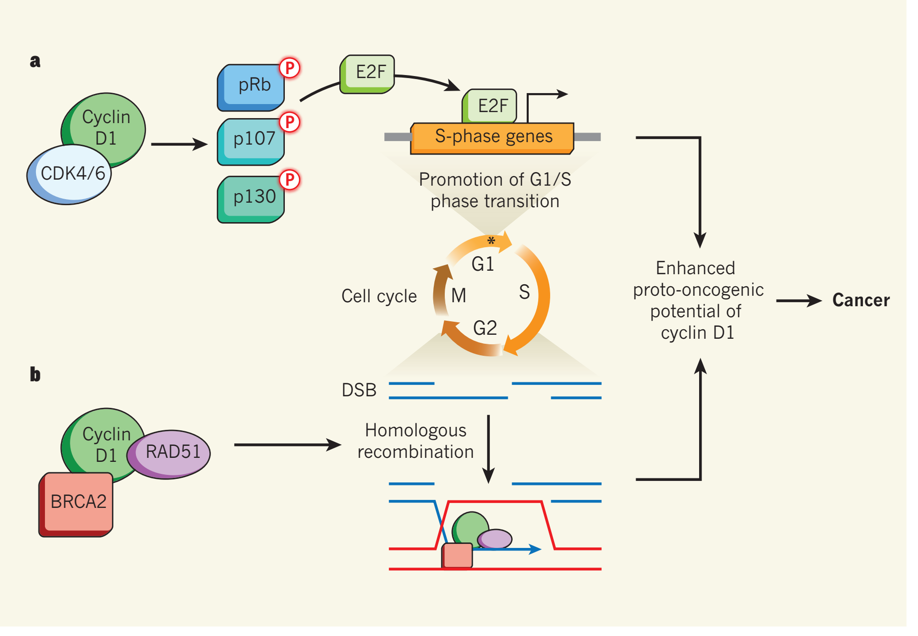


Figure 1 | Dual role of cyclin D1. a, As part of its function in the cell cycle, cyclin D1 interacts[2] with the enzymes CDK4 or CDK6, initiating phosphorylation (P) of the retinoblastoma family of transcriptional repressors (pRb, p107, p130). As a result, E2F transcription factors are liberated, allowing synthesis of S-phase genes[5]. This activity of cyclin D1–CDK4/6 peaks in the late G1 phase (asterisk) and is among the main determinants of the initiation of DNA replication and autonomous completion of the cell cycle[2,5]. **b**, Jirawatnotai *et al.*[3] report that cyclin D1 has another, non-catalytic role beyond the G1/S boundary. By directly binding to the RAD51 enzyme, cyclin D1 cooperates with BRCA2 to assemble and/or stabilize a central repair complex that acts on DNA double-strand breaks (DSBs) by homologous recombination. Both the cell-cycle and DNA-repair functions may contribute to the potential of cyclin D1 to become oncogenic when expressed at high levels in human tumours.

plethora of cellular proteins. Consistent with their role in cell proliferation, cyclins at abnormally high levels, particularly cyclins D1 and E, contribute to the development of diverse types of cancer[2].

To gain further insight into the function of cyclin D1, Jirawatnotai *et al.*[3] purified the protein from several types of human cancer cells and assessed the spectrum of proteins that interact with it. Surprisingly, several of the cyclin-D1 partner proteins turned out to be known components of the DNA-repair machinery. Particularly intriguing was the interaction between cyclin D1 and RAD51, which is a key enzyme involved in homologous recombination — a process that operates in the advanced S and G2 phases of the cell cycle to seal DNA breaks using the intact copy of the DNA sequence as a template[1,4]. The authors found that cyclin D1 binds to RAD51 directly, and that this interaction is stronger in cells exposed to ionizing radiation[3], despite the fact that this genotoxic insult leads to the partial degradation of cyclin D1.

Encouraged by these results, the authors depleted human cells of cyclin D1 and found not only that the cells became more sensitive to DNA-damaging agents such as ionizing radiation or various drugs, but also that they were slower to repair broken DNA and were defective in performing homologous recombination, although not other DNA-repair pathways. Reassuringly, the authors could remedy these defects by re-expressing normal cyclin D1 in such cells. However, a mutant form of the protein that cannot interact with RAD51 was ineffective.

As if the discovery of another function for cyclin D1 wasn't surprising enough, the authors further show that the role of cyclin D1 in DNA repair is independent of its CDK-associated activity[3]. This clearly separates the classical, CDK-dependent cell-cycle regulatory function of cyclin D1 from its role in DNA repair (Fig. 1). The different roles of cyclin D1 are relevant for cancer biology and treatment. For instance, Jirawatnotai *et al.* report that the DNA-repair function of this protein is required in cancer cells deficient in a tumour suppressor called retinoblastoma protein — a setting in which the cell-cycle-regulatory activity of cyclin D1 is dispensable[5] (Fig. 1).

How does cyclin D1 exert its role in DNA repair at a molecular level? Jirawatnotai *et al.*[3] show that this protein is localized to damaged chromosomal sites, facilitating recruitment of RAD51. They find that cyclin D1 itself is brought to such DNA breaks by BRCA2, another protein that is an established component of homologous recombination (Fig. 1).

Together these findings, which indicate that a core cell-cycle regulator has a non-catalytic function in DNA repair, raise a host of issues, both mechanistic and conceptual. For example, it is unclear how the relatively minor fraction of cyclin D1 that escapes degradation after DNA damage is modified or 'primed' to engage in homologous recombination. This is even more remarkable in the case of cancers involving retinoblastoma deficiency, which are known[5] to harbour generally lower levels of cyclin D1. This is not a purely academic issue, because Jirawatnotai and co-workers show that depleting cyclin D1 in retinoblastoma-deficient human tumours grown in mice may sensitize these cancers to radiotherapy. Another question is whether the related cyclins D2 and D3 also participate in the DNA-damage response, especially given that cyclin D1 is not expressed in some cell types.

Mechanistically, it seems unlikely that cyclin D1 is part of the core machinery responsible for homologous recombination. Yeast cells do not have D-type cyclins, and yet they rely on homologous recombination more than do vertebrates[4]. And the BRCA2-dependent loading of RAD51 at DNA lesions can be reconstituted in a test tube in the absence of cyclin D1[6,7]. Therefore, it is more likely that, in vertebrates, cyclin D1 has evolved to facilitate homologous recombination by simultaneously binding RAD51 and BRCA2, and so increasing their local concentration at damaged chromosomes.

Arguably most puzzling is the apparent discrepancy between the tumour-promoting, cell-cycle function of cyclin D1 and its role in DNA repair, which — by analogy with other repair processes[1,8] — should help to prevent cancer. Perhaps this conundrum would be resolved if cyclin D1 facilitated DNA repair, and hence survival of cancer cells, in the face of the severe, endogenous, oncogene-induced DNA damage to which the cells are exposed as well as the damage induced by genotoxic therapy[1,9]. Such a concept is plausible because, like cyclin D1, RAD51 is often expressed at high levels in tumours, and this might promote the fitness of such cancers and potentially increase their resistance to treatment. ■

Jiri Bartek *and* **Jiri Lukas** *are at the Danish Cancer Society, Centre for Genotoxic Stress Research, Copenhagen, Denmark. J.B. is also at the Institute of Molecular and Translational Medicine, Palacky University, Olomouc, Czech Republic.*
e-mails: jb@cancer.dk; jil@cancer.dk

1. Jackson, S. P. & Bartek, J. *Nature* **461,** 1071–1078 (2009).
2. Sherr, C. J. & Roberts, J. M. *Genes Dev.* **18,** 2699–2711 (2004).
3. Jirawatnotai, S. *et al. Nature* **474,** 230–234 (2011).
4. Mazón, G., Mimitou, E. P. & Symington, L. S. *Cell* **142,** 646 (2010).
5. Bartek, J. *et al. Curr. Opin. Cell Biol.* **8,** 805–814 (1996).
6. Jensen, R. B., Carreira, A. & Kowalczykowski, S. C. *Nature* **467,** 678–683 (2010).
7. Thorslund, T. *et al. Nature Struct. Mol. Biol.* **17,** 1263–1265 (2010).
8. Hoeijmakers, J. H. J. *Nature* **411,** 366–374 (2001).
9. Halazonetis, T. *et al. Science* **319,** 1352–1355 (2008).

doi:10.1038/nature10143

Role of the ubiquitin-like protein Hub1 in splice-site usage and alternative splicing

Shravan Kumar Mishra[1], Tim Ammon[1], Grzegorz M. Popowicz[2], Marcin Krajewski[2]†, Roland J. Nagel[3]†, Manuel Ares Jr[3], Tad A. Holak[2] & Stefan Jentsch[1]

Alternative splicing of pre-messenger RNAs diversifies gene products in eukaryotes and is guided by factors that enable spliceosomes to recognize particular splice sites. Here we report that alternative splicing of *Saccharomyces cerevisiae SRC1* pre-mRNA is promoted by the conserved ubiquitin-like protein Hub1. Structural and biochemical data show that Hub1 binds non-covalently to a conserved element termed HIND, which is present in the spliceosomal protein Snu66 in yeast and mammals, and Prp38 in plants. Hub1 binding mildly alters spliceosomal protein interactions and barely affects general splicing in *S. cerevisiae*. However, spliceosomes that lack Hub1, or are defective in Hub1–HIND interaction, cannot use certain non-canonical 5′ splice sites and are defective in alternative *SRC1* splicing. Hub1 confers alternative splicing not only when bound to HIND, but also when experimentally fused to Snu66, Prp38, or even the core splicing factor Prp8. Our study indicates a novel mechanism for splice site utilization that is guided by non-covalent modification of the spliceosome by an unconventional ubiquitin-like modifier.

Covalent modification of proteins by ubiquitin and related proteins (collectively called ubiquitin-like modifiers, UBLs) often critically alters substrate activity by influencing metabolic stability, binding behaviour or localization[1]. The switch-like properties of UBLs are crucial for pathways that regulate, for example, signal transduction, protein sorting, DNA repair, and development[1]. Covalent conjugation of a UBL to a substrate's target residue is ATP dependent, involves an enzyme cascade, and usually requires a free di-glycine (GG) motif at the protruding carboxy-terminal end of the UBL. Archetypal UBLs (ubiquitin, SUMO, Rub1 (also known as Nedd8)) are expressed as inactive precursors with C-terminal extensions. These extensions are removed by UBL-specific proteases, exposing the crucial C-terminal GG motif. Enzymes of this class also mediate UBL deconjugation, thus making the UBL-dependent switch reversible[1].

Hub1 (homologous to ubiquitin; known as UBL5 or beacon in mammals), another evolutionarily highly conserved UBL, is unique in lacking a protruding C-terminal tail with a GG motif. Instead, Hub1 possesses a C-terminal double tyrosine (YY) motif, followed by a non-conserved amino acid residue[2,3]. Although Hub1 from various organisms has been studied to some extent[4–8], its function remains poorly understood. Whereas *S. cerevisiae* cells deficient in Hub1 are viable and exhibit only minor phenotypes under normal growth conditions[6,7], the corresponding mutant of *Schizosaccharomyces pombe* is lethal[4,8]. One study reported that Hub1 forms covalent conjugates similar to ubiquitin and proposed that Hub1 is synthesized as a precursor and matured by processing C terminally of the YY motif[6]. However, no Hub1-specific processing, conjugation or deconjugation enzymes have been identified. Further studies have ruled out that Hub1 functions as a covalent protein modifier[7,8]; in fact, Hub1 was found to bind proteins non-covalently and independently of ATP, and the YY motif was shown to be nonessential[7,8].

Hub1 has been linked to various physiological functions, including cell cycle progression and polarized growth[6], the mitochondrial unfolded protein response[9], and mRNA splicing[4,8]. Conditional mutants of *S. pombe HUB1* show moderate RNA splicing defects, particularly at high temperatures, and Hub1 formed a non-covalent association with the spliceosomal (U4/U6.U5) tri-small nuclear ribonucleoprotein particle (snRNP) protein Snu66 (refs 4, 8). However, how the Hub1–Snu66 interaction affects splicing is unclear. It has been proposed that Hub1 is required for the nuclear localization of Snu66 (ref. 4), but Hub1 may affect the spliceosome directly and influence its activity.

Here we show that Hub1, through binding to Snu66, modifies the spliceosome in a way that enables it to tolerate and use certain non-canonical 5′ splice sites. We discovered that Hub1 binds Snu66 through an element called HIND in a unique, sequence-specific manner. We propose that Hub1 operationally resembles UBLs, with the important difference that Hub1 modifies substrates through non-covalent binding.

Hub1 binds to HINDs of spliceosomal proteins

Hub1 has been shown to bind the tri-snRNP protein Snu66 in yeast two-hybrid (Y2H) assays[4,10]. To verify this interaction *in vivo*, we raised antibodies specific for yeast Hub1 and Snu66 for co-immunoprecipitation assays. The antibodies against Hub1 bring down Snu66 from cell extracts and vice versa (Fig. 1a, b). Similar experiments with an altered form of Hub1 (YY changed to AA) showed that binding to Snu66 was independent of the YY motif (data not shown).

Amino-terminal fragments of Snu66 were both sufficient and required for Hub1 binding. Inspection of the protein sequence revealed two highly similar elements (72% identity) arranged in tandem (Fig. 1c, d). Fragments harbouring these repeats, either singly or in tandem, strongly interacted with Hub1 in Y2H assays (Supplementary Fig. 1a). Notably, the elements do not seem to bind ubiquitin or SUMO in Y2H assays, and Hub1 does not seem to bind classical ubiquitin-binding motifs like UBA or UIM (data not shown). The minimal polypeptide sequence defined by this assay was 18–19 amino acids long; because it had no obvious similarity to known motifs, we termed it HIND

[1]Department of Molecular Cell Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany. [2]NMR Spectroscopy, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany. [3]Center for Molecular Biology of RNA, Department of Molecular, Cell & Developmental Biology, University of California, Santa Cruz, California 95064, USA. †Present addresses: Institut für Biomedizinische Technik, ETH Zürich, Wolfgang-Pauli-Str. 10, 8093 Zürich, Switzerland (M.K.); Enzymology Research and Development, Life Technologies, 850 Lincoln Centre Drive, Foster City, California 94404, USA (R.J.N.).

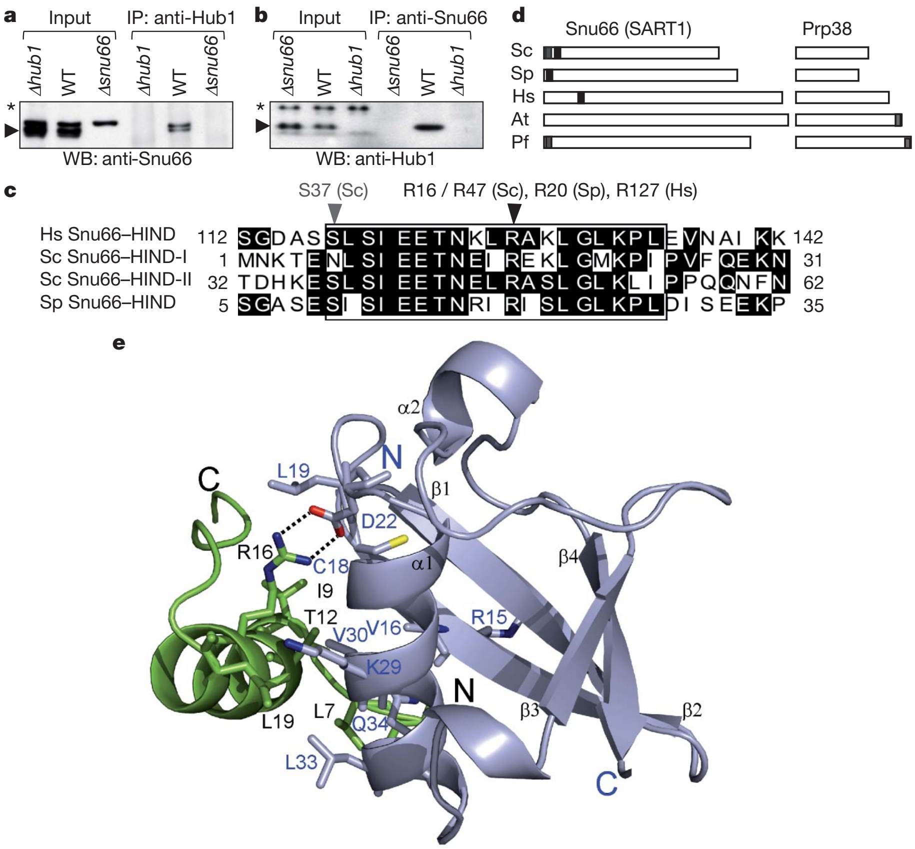


Figure 1 | **Hub1 binds to HIND.** **a**, Immunoprecipitation (IP) of Hub1 and co-immunoprecipitation of Snu66 from the indicated yeast strains. WB, western blot; WT, wild type. **b**, Immunoprecipitation of Snu66 and co-immunoprecipitation of Hub1 from the indicated yeast strains (asterisks denote cross-reacting signals). **c**, Clustal-W alignment of Snu66–HINDs from different organisms. Numbers give the positions of amino acids, and arrows indicate conserved Arg and Ser residues. At, *Arabidopsis thaliana*; Hs, *Homo sapiens*; Pf, *Plasmodium falciparum*; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*. **d**, Diagram of homologues of the spliceosomal proteins Snu66 (SART1) and Prp38, and the occurrence of HINDs (shaded). **e**, Crystal structure of Hub1 in complex with HIND-I peptide shown as ribbon plot. The interaction interface is formed by a salt bridge between Arg 16(47) (numbers in parentheses refer to HIND-II) of Snu66 and Asp 22 of Hub1, accompanied by a patch of hydrophobic interactions. The hydrophobic residues of HINDs that participate in the Hub1-binding interface are Leu 7(38), Ile 9(40), Thr 12(43), Ile 15 (Leu 46), Arg 16(47) (Cβ and Cγ), Leu 19(50), Met 21 (Leu 52) and Ile 24(55). On Hub1, the interface is formed by Leu 19, Val 30, and additionally Met 1, Val 16, Lys 17 (Cβ, Cγ, Cδ), Cys 18, Lys 29 (Cβ, Cγ, Cδ) and Leu 33.

(Hub1-interaction domain). GST pull-down assays with purified proteins confirmed that Hub1–Snu66 binding is direct and mediated via HIND (Supplementary Fig. 1b, c). Gel filtration analysis indicates that the N-terminal domain of Snu66 can bind two Hub1 molecules (Supplementary Fig. 1d).

Snu66 proteins (also known as SART1) from Saccharomycotina (for example, *S. cerevisiae*, *Candida*) and some Stramenopiles possess two N-terminal HINDs, but the homologues from *S. pombe* and vertebrates possess only one HIND at this position (Fig. 1d). As Hub1 binds Snu66 via its HIND in *S. pombe* and in humans also (Supplementary Fig. 2a–e), the mechanism of Hub1 recruitment seems to be conserved. Intriguingly, plant Snu66 homologues (and also for example, Amoebazoa) lack HIND sequences; this absence is compensated by HINDs found in C-terminal extra domains of proteins related to the spliceosomal protein Prp38 (Fig. 1d and Supplementary Fig. 2f). Furthermore, in *Plasmodium*, functional HINDs are present in both Snu66 and Prp38 homologues, as they bind Hub1 and involve the same surface interface (Supplementary Fig. 2g, h). Because in yeast and mammalian cells Snu66 and Prp38 are constituents of the spliceosome[11], HINDs may be evolutionarily associated (and perhaps restricted) to splicing proteins. Moreover, as the identity of the HIND-bearing spliceosomal protein seems to be flexible, HINDs may function irrespective of their exact positioning within the spliceosome (see later). By extension, this observation indicates that Hub1 may affect spliceosome activity as a whole, rather than functionally modulate the protein it directly binds. As Hub1 colocalizes with Snu66 in nuclear speckles in human cells (Supplementary Fig. 3a), it seems likely that the function of Hub1 in splicing is conserved.

Structure of Hub1 in complex with HIND

To determine whether Hub1 functions as a non-covalently acting modifier by interacting with HIND, we sought to gain molecular insights into the Hub1–HIND interaction. The characterization of peptides corresponding to the two HINDs of *S. cerevisiae* Snu66 (HIND-I, 18 amino acids; HIND-II, 19 amino acids) by circular dichroism (CD) and NMR revealed that an isolated HIND peptide is apparently helical in solution (Supplementary Fig. 4a, b), an unusual feature for such short peptides.

Next, we solved the structure (1.4 Å) of Hub1 in complex with each Snu66 HIND peptide. The structure of Hub1 is highly similar to ubiquitin[2,3,12] (PDB code 1UBI; root mean squared deviation (r.m.s.d.) of 0.88 Å for the main chain heavy atoms) and SUMO[13] (PDB code 1A5R; r.m.s.d. of 1.72 Å), and consists of a half-open β barrel completed with two flanking α helices; the secondary structure elements have a ββαβαβ pattern (Fig. 1e, Supplementary Figs 5, 6 and Supplementary Table 2).

The observed interactions of Hub1 with the Snu66 HIND-I and HIND-II peptides are almost identical (Fig. 1e and Supplementary Fig. 5a). The bound peptide forms an 11-residue helix (Ile 3–Leu 13,

numbering of HIND-I) with its C-terminal part flipped over along the peptide helix (Fig. 1e and Supplementary Fig. 5a). The whole Hub1–HIND interface has a surface of about 500 Å^2. On HIND binding, the main-chain fold of Hub1 does not change, but a number of side chains are significantly affected (Supplementary Fig. 5b). The Hub1–Snu66 interactions seen in the X-ray structures were fully corroborated by our mutational studies. For example, replacement of the salt-bridge-forming residues Asp 22 of Hub1 and Arg 16(47) (number in parentheses refers to residues in HIND-II) of Snu66 by Ala abolished Hub1–Snu66 binding (Supplementary Fig. 5c–f).

Most unexpectedly, the structure of Hub1–HIND revealed a new binding paradigm unseen in interactions of ubiquitin and ubiquitin-like proteins with their binding partners. Most ubiquitin-binding modules of ubiquitin receptors bind to the hydrophobic surface of ubiquitin centred on Ile 44 (ref. 14), which is almost exactly on the site opposing the HIND-binding face of the ubiquitin fold (Supplementary Fig. 6a). The ubiquitin interactions with ubiquitin receptors have mostly hydrophobic character and are usually weak ($K_d \geq 100\ \mu$M), whereas the Hub1–HIND interaction comprises a strong salt bridge accompanied by several hydrophobic contacts and high affinity (K_d Hub1–Snu66: $0.59 \pm 0.07\ \mu$M; K_d Hub1–HIND-I: $1.69 \pm 0.27\ \mu$M). The mode of interaction of Hub1–HIND is also clearly different from that seen for SUMO interacting with SUMO-interaction motifs (SIM)[15], although SIMs bind to a similar surface of the ubiquitin fold (Supplementary Fig. 6b).

By performing heteronuclear single quantum coherence (HSQC) NMR, we found that $^{15}N^2H$-labelled Snu66, including its HIND-bearing N-terminal domain, is unstructured (Supplementary Fig. 7a, b). As co-expression of Hub1 significantly increased the solubility of Snu66 in bacteria (Supplementary Fig. 7c), we speculated that Hub1 might influence Snu66 folding. Indeed, when unlabelled Hub1 was titrated into $^{15}N^2H$-labelled Snu66, several distinct peaks appeared, indicating that a part of the protein acquires structure after Hub1 binding (Supplementary Fig. 7a, b). Comparison of an N-terminal 65-amino-acid fragment and full-length Snu66 revealed that Hub1-induced folding is restricted to the HIND-containing N terminus of Snu66. This finding further corroborates the idea that Hub1 modifies the spliceosome rather than modulating the properties of an individual binding partner.

Hub1 modifies the spliceosome

Snu66 is a conserved component of the spliceosomal tri-snRNP complex[16,17]. To investigate whether Hub1 alters the composition of the spliceosome, we immunoprecipitated Snu66 from wild-type, *Δhub1* or *HUB1*-overexpressing strains (*Δsnu66* strain as control), and identified co-purifying proteins by Orbitrap mass spectrometry (Supplementary Fig. 8a–c). The set of identified proteins in all three samples significantly overlapped with the catalogue of proteins in the yeast spliceosomal complex B (penta-snRNP)[18,19]. Although we did not observe major changes in the protein composition of the three samples, in Hub1-deficient cells we reproducibly found an overrepresentation of certain proteins from U1 and U2 snRNPs, but not of the tri-snRNP (Supplementary Fig. 8a–d); by contrast, no significant spliceosomal alterations were observed in the strain overexpressing Hub1. We confirmed these findings by Snu66-directed immunoprecipitation using strains that express tagged versions of selected proteins from the tri-, U1-, and U2-snRNPs from their genomic loci (Supplementary Fig. 8d). We conclude that modification of the spliceosome by Hub1 only moderately affects spliceosome composition, so that the basic makeup of the snRNPs is preserved. Hub1 does bind spliceosomes *in vivo*, as we could co-immunoprecipitate the central spliceosomal protein Prp8 with Hub1 antibodies, in a reaction that is mediated by Snu66 (Supplementary Fig. 8e).

Functional links to the spliceosome

Hub1 is present in both cytosol and nucleus in yeasts and mammals, whereas Snu66 appears to be mostly nuclear (Supplementary Fig. 3a, b)[4,8]. It was reported that Hub1 is required for the nuclear localization of Snu66 (ref. 4), but we found that Snu66 is nuclear even in *S cerevisiae Δhub1* cells, and that a Snu66 variant deficient in Hub1 binding is nuclear in *S. cerevisiae*, *S. pombe* and human cells (Supplementary Fig. 3a, b and data not shown). Although in *S. pombe HUB1* and *SNU66* are essential[4,8], mutants deficient in Hub1–Snu66 interaction are viable (Supplementary Fig. 10a, b). This finding, together with the detection of significant cytosolic pools of Hub1 and the observation that Hub1 is apparently more abundant in cells than Snu66 (data not shown), indicates that the function of the Hub1 modifier may not be restricted to splicing.

We observed synthetic sickness or lethality of *Δhub1* and *Δsnu66* mutants if they were combined with mutant alleles of a number of spliceosomal genes (Supplementary Fig. 9a, b). Notably, the phenotype of *Δhub1 prp8** (*prp8** refers to *prp8(P1384L)*) could not be rescued by expressing *hub1(D22A)*, which encodes a Hub1 variant defective in Snu66 interaction (Fig. 2a). Similarly, the lethal *Δsnu66 prp8** double mutant could not be complemented by expression of Snu66 variants that lack the two HINDs (Fig. 2b). Thus, we conclude that Hub1, Snu66 and the Hub1–HIND interaction are indeed relevant for splicing.

As noted earlier, HIND elements in *S. cerevisiae* are located within the N-terminal domain of Snu66, but in plants, HIND is present as a C-terminal extension of the spliceosomal protein Prp38 (Fig. 1d). To address whether the specific localization of HIND is important, we fused a HIND element to the C terminus of yeast Prp38 (*PRP38–HIND*) and determined functionality as above. We found that *PRP38–HIND* could indeed functionally rescue *snu66(ΔHIND)* in the synthetic lethality assay with *prp8**, and that this activity depends on HIND residues crucial for Hub1 binding (Supplementary Fig. 9d). This finding confirms a remarkable plasticity of the Hub1 modifier, in that the exact positioning of HIND (and thereby of Hub1), on the spliceosome is not critical for function.

The *S. cerevisiae Δhub1 Δsnu66* double—but not the corresponding single—mutant is temperature sensitive (Supplementary Fig. 9b). This non-epistatic behaviour suggests partially separate functions and may again point to roles of Hub1 in addition to splicing. Whereas the removal of the YY motif does not interfere with Hub1 activity[4,7,8], Hub1 variants possessing certain charged C-terminal extensions failed to support the vital activity of Hub1 in *S. pombe* (Supplementary Fig. 10c), although the proteins can still bind HIND (Supplementary Fig. 10d). Moreover, also in *S. cerevisiae*, expression of a Hub1 variant with a short extension (*hub1–DD*) could not suppress the synthetic sickness of *Δhub1 prp8** (Fig. 2a). Thus, in addition to the HIND-binding surface the area neighbouring the C terminus of Hub1 is functionally important, perhaps for additional physical interactions.

To investigate the splicing competence of Hub1-deficient cells directly, we used a splicing-sensitive microarray that can distinguish spliced from unspliced RNAs of almost all intron-containing *S. cerevisiae* genes[20,21]. Onto this array we hybridized RNA samples from wild-type, *Δhub1*, *Δsnu66* and *Δhub1 Δsnu66* cultures (Fig. 2c). We found virtually no splicing defects in the *Δhub1* mutant for all transcripts, with the only discernable exception of *RPL34B*, the splicing of which was mildly affected. On the other hand, we noticed small splicing defects for *Δsnu66*, which were aggravated if the cells were additionally defective in Hub1 (*Δhub1 Δsnu66*) (Fig. 2c and Supplementary Table 3). To confirm these data, we isolated RNA from these strains and analysed splicing of a selected set of RNAs by quantitative polymerase chain reaction with reverse transcription (RT–qPCR). Again, splicing was virtually normal in *Δhub1* cells, and only a small splicing defect was discernable for *RPL34B* (Supplementary Fig. 11a). However, this small defect in *RPL34B* splicing is rather unlikely to affect the level of this ribosomal protein, as splicing of *RPL34A*, the gene product of which is almost identical to that of *RPL34B*, is not affected (data not shown). We found that *HUB1* transcripts are inducible by cadmium and on oxidative stress (Supplementary Fig. 11b),

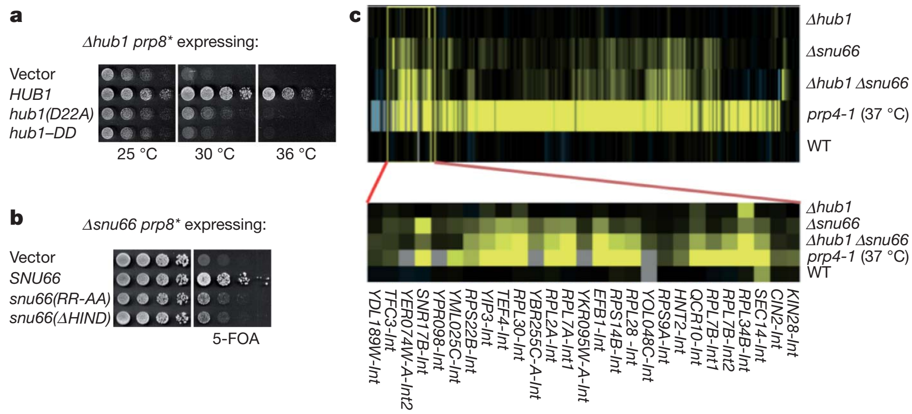


Figure 2 | Functional links to splicing. **a**, Rescue of synthetic sickness of *Δhub1 prp8** by expression of *HUB1* wild type, but not *hub1(D22A)* and *hub1–DD*. **b**, Rescue of synthetic lethality of *Δsnu66 prp8** by expression of *SNU66* wild type, but not by *snu66(RR-AA)* or *snu66(ΔHIND)*. Wild-type *SNU66* expressed from a *URA3*-containing plasmid was shuffled out by counter-selection on 5-fluoroorotic acid (5-FOA) plates. **c**, Microarray analysis for splicing defects. Diagram of intron accumulation (yellow) in strains deleted of *HUB1* and *SNU66* (data in Supplementary Table 3). The top panel shows relative splicing efficiency for almost all yeast introns as measured by splicing sensitive microarrays[20,21]. In each case RNA is derived from the indicated mutant, and wild-type RNA (comparing two different RNA preparations from wild-type cells) is used for reference. The temperature-sensitive splicing mutant *prp4-1* is the positive control. The bottom panel is an expanded view of a set of transcripts that are particularly affected in *Δhub1 Δsnu66* and also includes an intron of *RPL34B* that is mildly affected in *Δhub1*.

but we detected no significant *Δhub1*-specific splicing defects under these (or heat shock) conditions (Supplementary Fig. 11c, d). Thus, Hub1 does not seem to significantly affect general splicing, and splicing defects only occur if the spliceosome has additional deficiencies. This indicates that Hub1 might have a silent (and redundant) role in general splicing, or that Hub1 fulfils a specific, splicing-related function.

Splice-site usage and alternative splicing

Splice sites in *S. cerevisiae* show very little sequence variation, and alterations in the sequence can affect splicing[22]. To address whether Hub1 affects splice-site usage, we used a reporter assay based on an intron-containing *RP51*–LacZ* fusion[23] (Fig. 3a) monitored by both RT–PCR and β-galactosidase (LacZ) activity. For the consensus 5′ splice site (GUAUGU), again we observed no splicing defects for *Δhub1*, but defects for *Δsnu66*, *prp8** and *Δhub1 prp8** (Fig. 3b). Intriguingly, when we altered the 5′-splice-site sequence in the reporter construct, some 5′-splice-site variants were similarly used in both wild-type and *Δhub1* cells, but certain others required specifically the presence of Hub1 for proper splicing (Fig. 3a, b and data not shown). In fact, splicing via the mutant 5′ splice site GUAUAU was nearly as defective in *Δhub1* as in *prp8**, and was almost abolished in *Δhub1 prp8** (Fig. 3b). Importantly, this splicing defect also occurred in strains that only express a Hub1 defective in Snu66 binding (*hub1(D22A)*), or Hub1 harbouring an abnormal C terminus (*hub1–DD*) (Fig. 3c). To corroborate and extend these findings we also used another splicing sensitive *ACT1–CUP1* reporter assay[24]. Again we found splicing defects of *Δhub1* cells for certain 5′-splice-site alterations (GUAUAU, GUCUGU), but not significantly for variants of the 3′ splice site (G/AG, U/UG instead of U/AG; slash symbolizes the intron/exon boundary) or the branch point (UCCUAAC, UACUACC instead of UACUAAC) (Supplementary Fig. 12).

Most metazoans exhibit a high variation of 5′ splice sites, and this plasticity is exploited for alternative splicing[22,25]. Regulation of these events seems to be guided by factors that enable the spliceosome to recognize divergent 5′ splice sites. In *S. cerevisiae*, in which 5′ splice sites are nearly invariant, alternative splicing is extremely rare[22]. In this organism, *SRC1* is the only known case of alternative splicing leading to two different proteins with different functions[26–28]. In both cases, a single (130- or 126-nucleotide long) intron is excised, yet by using two different, overlapping 5′ splice sites. Notably, both 5′ splice sites differ significantly from the consensus (GCAAGU, GUGAGU).

SRC1 was present on the splicing array we used, but the two splice forms could not be individually identified because of their overlapping

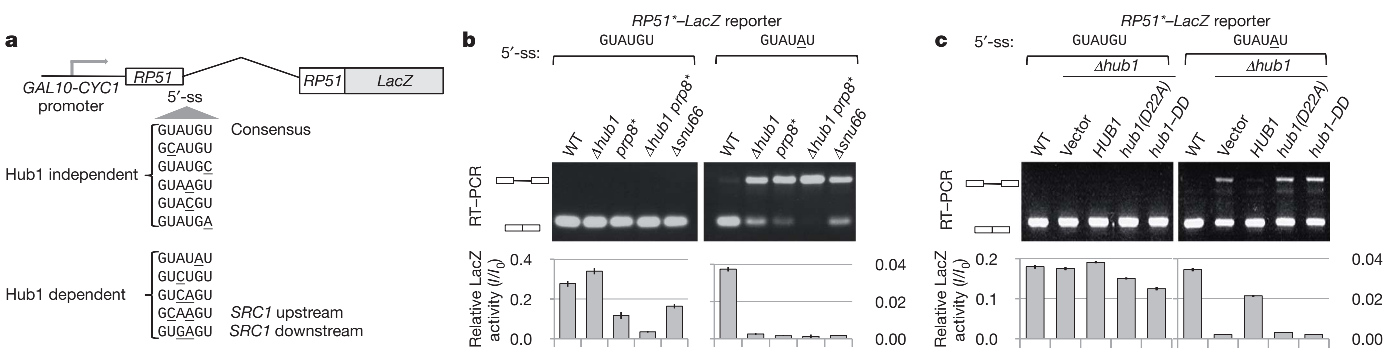


Figure 3 | Splice-site usage. **a**, Scheme for the requirement of Hub1 for efficient utilization of canonical and non-canonical 5′ splice sites (5′-ss; underlines mark changes) inserted into *RP51*–LacZ* fusions[23]. The individual 5′ splice sites of *SRC1* in the context of the reporter assay were used. **b**, **c**, Splicing defects in wild type and mutants with canonical (GUAUGU) and non-canonical (GUAUAU) 5′ splice sites. RT–PCR assay of total RNA isolated from yeast expressing different version of intron-containing *RP51*–LacZ* fusions. Bottom panels show ratios of corresponding LacZ (β-galactosidase) activities (I) to the activity from an intron-less (I_0) construct. Error bars show standard deviation of three samples. Strains in **c** are *Δhub1* and complemented by *HUB1* alleles.

5′ splice sites. When we sequenced the complementary DNA across the exon/exon boundaries of spliced *SRC1*, however, we observed a sequence mixture corresponding to the two splice forms in wild-type cells (and also in *Δsnu66* and *prp8** mutants), but largely only one form (corresponding to splicing via the downstream 5′ splice site) in *Δhub1* mutants (Fig. 4a and Supplementary Fig. 13a, b). To corroborate this finding, we made use of a chromosomally expressed TAP-tagged Src1 variant monitoring the expression of the two splice variants[28]. Indeed, splicing via the upstream 5′ splice site (GCAAGU) was again almost completely absent in Hub1-deficient cells, whereas splicing via the downstream 5′ splice site was normal or even more pronounced (Fig. 4b). Moreover, splicing via the upstream 5′ splice site was equally defective in cells that express Hub1 defective in HIND interaction (*hub1(D22A)*), whereas both forms were reduced in *Δsnu66*

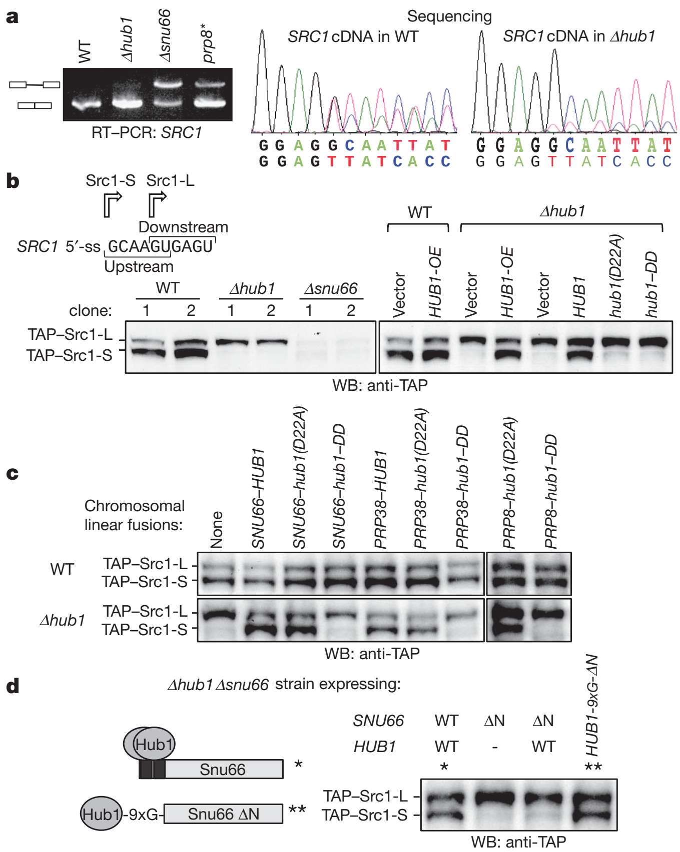


Figure 4 | Alternative splicing of *SRC1*. **a**, RT–PCR assay of *SRC1* transcripts in different mutants (left panel), and sequencing of *SRC1* cDNA from wild-type (WT) and *Δhub1* strains (right panels). Only general splicing defects of *Δsnu66* and *prp8** mutants are detectable by RT–PCR as the 5′ splice sites of *SRC1* are overlapping, but alternative splicing defects via the upstream 5′ splice site in the *Δhub1* strain are revealed by sequencing. **b**, Chromosomal *SRC1* was N-terminally TAP tagged to detect the two gene products (TAP–Src1-L, TAP–Src1-S) generated by alternative splicing. Splicing via the upstream (GCAAGU) and downstream (GUGAGU) 5′ splice sites (5′-ss) is indicated (top left). Alternative splicing defects (usage of the upstream 5′ splice site) in *Δhub1* and general splicing defects in *Δsnu66* were detected by western blotting using TAP-tag-specific antibodies. Wild-type and mutant strains, and *Δhub1* strains expressing Hub1 variants, or which overexpress *HUB1*, were used. **c**, *SRC1* alternative splicing is restored in *Δhub1* harbouring chromosomally expressed linear fusions of *PRP8*, *PRP38* and *SNU66* genes with *HUB1* or *hub1(D22A)* but not with *hub1–DD* (assay similar to **b**). **d**, *SRC1* alternative splicing is supported by a linear fusion of Hub1 with Snu66ΔN in a *Δhub1 Δsnu66* strain. The *Δhub1 Δsnu66* strain with TAP-tagged Src1 was transformed with plasmids expressing *HUB1*, *SNU66*, or a construct encoding Hub1 fused to the N terminus of the Snu66 variant that lacks the HIND region (ΔN), separated by a poly-glycine (9×G) linker. Asterisks indicate the two ways of Hub1–Snu66 complex formation.

(Fig. 4b). We also detected both forms of the Src1 protein in *Δprp17* and *prp8** mutants (Supplementary Fig. 13c), indicating that the choice of *SRC1* 5′ splice site is not significantly altered in these general splicing mutants. Notably, we could make *SRC1* splicing via the upstream 5′ splice site independent of Hub1 by mutating this 5′ splice site, but this caused a repression of splicing via the downstream 5′ splice site (Supplementary Fig. 13d). Conversely, certain mutations in the downstream 5′ splice site strongly affected splicing via the upstream 5′ splice site even when Hub1 was present (wild-type cells). We also tested the two *SRC1* 5′ splice sites individually in the *RP51**–*LacZ* reporter and found that in isolation, both non-canonical elements require Hub1 for full splicing activity (Fig. 3a; data not shown). This indicates that the characteristic differential Hub1 dependence of *SRC1* alternative splicing requires the tandem arrangement of overlapping 5′ splice sites. In this context it is interesting to note that alternative *SRC1* splicing might be regulated, as we found that specifically the Hub1-dependent splicing product is enriched in the G2 phase of the cell cycle (Supplementary Fig. 13e).

Compared to *S. cerevisiae*, *S. pombe* possesses a much higher number of intron-containing genes, and its splicing machinery resembles more closely its mammalian counterpart[22,29]. Similarly, the sequences of *S. pombe* 5′ splice sites are highly divergent and several cases of alternative splicing are known[22,30]. To address whether Hub1 also mediates selective splicing in *S. pombe* we focused on two different pre-mRNAs. We tested *CDC2* (encoding cyclin-dependent kinase), as it possesses four introns with different 5′ splice sites. We found that splicing of intron 3 in particular, which possesses a rarely used 5′ splice site (GUAUAU; Hub1-dependent in *S. cerevisiae*; Fig. 3a), needs Hub1 for normal splicing (Supplementary Fig. 14a–d). *ZAS1* (encoding a transcription factor), on the other hand, is known to be alternatively spliced as sometimes intron 2 is retained, yielding a larger open reading frame[30]. Again we found that Hub1 is needed for normal *ZAS1* splicing and that intron inclusion occurs in *hub1* mutants at a much higher frequency (Supplementary Fig. 14d). Thus, Hub1-guided selective splicing is a conserved mechanism, and may occur in all eukaryotes.

Using the *SRC1* splicing assay, we examined the significance of our initial observation that HINDs are present on Snu66, Prp38 or both spliceosomal proteins in different organisms. Remarkably, alternative *SRC1* splicing was normal in *Δhub1 S. cerevisiae* cells in which linear (C-terminal) fusions of either Snu66 or Prp38 with Hub1 were chromosomally expressed (Fig. 4c). *SRC1* splicing was defective when the Hub1 moiety of the fusions was C-terminally altered (Hub1–DD), but fully functional when the Hub1 moiety was deficient in Snu66 interaction. Hub1 fused to Snu66ΔN (lacking the HINDs) via a polyglycine linker may also support Hub1-dependent alternative *SRC1* splicing even if free Hub1 and wild-type Snu66 are absent (Fig. 4d). Surprisingly, when we fused HIND-binding defective Hub1 (Hub1(D22A)) to the core spliceosomal protein Prp8, alternative *SRC1* splicing was supported (Fig. 4c). These findings indicate that linear Hub1 fusions functionally mimic HIND-mediated Hub1 recruitments, and show that Hub1-dependent effects on splicing act independently of a specific localization on the spliceosome.

Conclusions

Alternative splicing substantially increases the gene product repertoire and is a major source of cell type differentiation[22,25]. Metazoans in particular use this mechanism extensively, and it is estimated that the majority of human pre-mRNAs undergo alternative splicing[22]. Conventional alternative splicing is largely controlled by positively acting SR (Ser-Arg) proteins and negatively acting heterogeneous nuclear ribonucleoprotein particles (hnRNPs).

We have discovered a new principle for splice-site utilization, which involves non-covalent binding of the ubiquitin-related protein Hub1 to the spliceosome. Notably, Hub1 in *S. cerevisiae* is not required for general splicing and the usage of canonical 5′ splice sites

(Fig. 2c), but is required for the usage of certain non-canonical 5′ splice sites. Hub1 might function as a 'plasticity factor', which, by relaxing the spliceosome's specificity, enables it to act productively on divergent 5′ splice sites. Given the high conservation of Hub1 and HIND, and the fact that Hub1-dependent splicing operates through evolutionary divergent spliceosomes, it seems highly likely that Hub1-controlled splicing occurs universally in eukaryotes. SR proteins and hnRNPs involved in spliceosome targeting do not seem to exist in *S. cerevisiae*[22,31], and thus the Hub1-dependent mechanism may be evolutionarily older. It seems plausible that the more elaborate SR/hnRNP-guided mechanism may have co-evolved with the rise in gene complexity to generate multiple different gene products from a single gene.

Hub1 is structurally very similar to ubiquitin and equally highly conserved and ancient, yet the two proteins function in completely different ways. Hub1 binds proteins only non-covalently, and at least for splicing, Hub1 seems to functionally alter the complex it binds to, rather than modulating its direct binding partner. Because of this distinctive property, it is attractive to speculate that Hub1 might alter the structure of the spliceosome or provide novel binding surfaces for physical interactions.

METHODS SUMMARY

Yeast strains are listed in Supplementary Table 1. Protein methods, mass spectrometric analysis, interaction studies, splicing assays and microscopy are standard techniques and are described in Methods. NMR spectroscopy and X-ray crystallography are detailed in Methods, and the data collection and refinement statistics are summarized in Supplementary Table 2.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 November 2010; accepted 20 April 2011.

Published online 25 May; corrected 9 June 2011 (see full-text HTML version for details).

1. Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **458,** 422–429 (2009).
2. McNally, T. *et al.* Structural analysis of UBL5, a novel ubiquitin-like modifier. *Protein Sci.* **12,** 1562–1566 (2003).
3. Ramelot, T. A. *et al.* Solution structure of the yeast ubiquitin-like modifier protein Hub1. *J. Struct. Funct. Genomics* **4,** 25–30 (2003).
4. Wilkinson, C. R. *et al.* Ubiquitin-like protein Hub1 is required for pre-mRNA splicing and localization of an essential splicing factor in fission yeast. *Curr. Biol.* **14,** 2283–2288 (2004).
5. Friedman, J. S., Koop, B. F., Raymond, V. & Walter, M. A. Isolation of a ubiquitin-like (*UBL5*) gene from a screen identifying highly expressed and conserved iris genes. *Genomics* **71,** 252–255 (2001).
6. Dittmar, G. A., Wilkinson, C. R., Jedrzejewski, P. T. & Finley, D. Role of a ubiquitin-like modification in polarized morphogenesis. *Science* **295,** 2442–2446 (2002).
7. Lüders, J., Pyrowolakis, G. & Jentsch, S. The ubiquitin-like protein HUB1 forms SDS-resistant complexes with cellular proteins in the absence of ATP. *EMBO Rep.* **4,** 1169–1174 (2003).
8. Yashiroda, H. & Tanaka, K. Hub1 is an essential ubiquitin-like protein without functioning as a typical modifier in fission yeast. *Genes Cells* **9,** 1189–1197 (2004).
9. Benedetti, C., Haynes, C. M., Yang, Y., Harding, H. P. & Ron, D. Ubiquitin-like protein 5 positively regulates chaperone gene expression in the mitochondrial unfolded protein response. *Genetics* **174,** 229–239 (2006).
10. Hazbun, T. R. *et al.* Assigning function to yeast proteins by integration of technologies. *Mol. Cell* **12,** 1353–1365 (2003).
11. Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136,** 701–718 (2009).
12. Ramage, R. *et al.* Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem. J.* **299,** 151–158 (1994).
13. Bayer, P. *et al.* Structure determination of the small ubiquitin-related modifier SUMO-1. *J. Mol. Biol.* **280,** 275–286 (1998).
14. Dikic, I., Wakatsuki, S. & Walters, K. J. Ubiquitin-binding domains—from structures to functions. *Nature Rev. Mol. Cell Biol.* **10,** 659–671 (2009).
15. Song, J., Zhang, Z., Hu, W. & Chen, Y. Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. *J. Biol. Chem.* **280,** 40122–40129 (2005).
16. Makarova, O. V., Makarov, E. M. & Luhrmann, R. The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *EMBO J.* **20,** 2553–2563 (2001).
17. Stevens, S. W. & Abelson, J. Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc. Natl Acad. Sci. USA* **96,** 7226–7231 (1999).
18. Stevens, S. W. *et al.* Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol. Cell* **9,** 31–44 (2002).
19. Fabrizio, P. *et al.* The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* **36,** 593–608 (2009).
20. Clark, T. A., Sugnet, C. W. & Ares, M. Jr. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* **296,** 907–910 (2002).
21. Burckin, T. *et al.* Exploring functional relationships between components of the gene expression machinery. *Nature Struct. Mol. Biol.* **12,** 175–182 (2005).
22. Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: diversification, exon definition and function. *Nature Rev. Genet.* **11,** 345–355 (2010).
23. Jacquier, A., Rodriguez, J. R. & Rosbash, M. A quantitative analysis of the effects of 5′ junction and TACTAAC box mutants and mutant combinations on yeast mRNA splicing. *Cell* **43,** 423–430 (1985).
24. Lesser, C. F. & Guthrie, C. Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, *CUP1*. *Genetics* **133,** 851–863 (1993).
25. Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Rev. Mol. Cell Biol.* **10,** 741–754 (2009).
26. Davis, C. A., Grate, L., Spingola, M. & Ares, M. Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res.* **28,** 1700–1706 (2000).
27. Rodríguez-Navarro, S., Igual, J. C. & Pérez-Ortín, J. E. *SRC1*: an intron-containing yeast gene involved in sister chromatid segregation. *Yeast* **19,** 43–54 (2002).
28. Grund, S. E. *et al.* The inner nuclear membrane protein Src1 associates with subtelomeric genes and alters their regulated gene expression. *J. Cell Biol.* **182,** 897–910 (2008).
29. Kuhn, A. N. & Kaufer, N. F. Pre-mRNA splicing in *Schizosaccharomyces pombe*: regulatory role of a kinase conserved from fission yeast to mammals. *Curr. Genet.* **42,** 241–251 (2003).
30. Okazaki, K. & Niwa, O. mRNAs encoding zinc finger protein isoforms are expressed by alternative splicing of an in-frame intron in fission yeast. *DNA Res.* **7,** 27–30 (2000).
31. Kress, T. L., Krogan, N. J. & Guthrie, C. A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. *Mol. Cell* **32,** 727–734 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank U. Cramer and M. Kost for technical assistance, G. Dittmar, C. Guthrie, M. Konarska, K. Matuschewski, O. Nielsen, M. Rosbash and H. Yashiroda for materials, S. Uebel and C. Boulegue for mass spectrometric analysis and help, K. Hofmann for pointing out putative HIND elements in Prp38 proteins of plants, and M. Singh for initiating structural work. S.J. is supported by the Max Planck Society, Deutsche Forschungsgemeinschaft, Fonds der chemischen Industrie, Center for Integrated Protein Science Munich and RUBICON EU Network of Excellence; T.A.H. by the Max Planck Society; M.A. by NIH (GM040478).

Author Contributions S.K.M. (*S. cerevisiae*, *S. pombe*), T.A. (mammalian) and S.J. designed, obtained and analysed the genetic, biochemical and functional data; G.M.P, M.K. and T.A.H. the structural data; R.J.N. and M.A. Jr the splicing array data. S.J. and S.K.M. wrote the paper, and all authors contributed to the manuscript.

Author Information Coordinates and the experimental structural factors of both complexes have been deposited in the PDB under the following codes: Hub1–HIND-I, 3PLU; Hub1–HIND-II, 3PLV. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.J. (Jentsch@biochem.mpg.de).

METHODS

Yeast strains, plasmids and DNA techniques. *S. cerevisiae* and *S. pombe* strains are listed in Supplementary Table 1. Yeast growth assays were performed by spotting fivefold serial dilutions of indicated strains on solid agar plates. Yeast strains isogenic to DF5, W303 and BY4741 were used for biochemical and genetic studies, and PJ69-7a for Y2H assays. All tagged strains and deletion mutants were constructed by a PCR-based strategy[32,33] and confirmed by western blot analysis and PCR, respectively. The *prp8** allele (*prp8(P1384L)*) was provided by G. Dittmar. For complementation studies, the complete *HUB1* and *SNU66* open reading frame (ORF), 1,000 bp of the upstream promoter and 500 bp of its terminator were cloned into centromeric plasmids. For Hub1 overexpression, the *HUB1* ORF was cloned into an integrative plasmid with the *TEF2* promoter and *ADH1* terminator. To generate 5′-splice-site variants for the *SRC1* gene, the full-length ORF was cloned into a centromeric plasmid downstream of the *GAL1-10* promoter. To generate the Hub1–Snu66ΔN linear fusion, DNA sequences encoding Hub1 plus a linker of nine glycine residues was inserted downstream of the *SNU66* promoter and upstream of the *SNU66* ORF. The vectors *pET28a-c* or *pGEX-5X1* were used for expression of 6×His-tagged or GST-tagged proteins, respectively, and *pGBDUC1* and *pGADC1* for Y2H assays. Plasmids with point mutations were constructed by site-directed mutagenesis using specific primers. *Plasmodium falciparum* genomic DNA (gift from K. Matuschewski) or a cDNA library of *Arabidopsis thaliana* (BioChain) were used as templates to generate the corresponding clones for interaction studies. For western-blot-based *SRC1* alternative splicing assays, a DNA sequence covering the *CUP1-1* promoter and TAP tag was integrated upstream of *SRC1* at its genomic loci. The *CUP1-1* promoter was induced for 4 h in the presence of 200 μM $CuSO_4$. For the construction of linear fusions of the splicing factors Snu66, Prp38 and Prp8 with Hub1 variants, DNA encoding *HUB1*, *hub1(D22A)* and *hub1–DD* were cloned into vector *pFA6a-natNT2*. The C-terminal fusions were finally generated by a PCR-based strategy[32,33]. For the (*GAL1-10*-promoter-driven) splicing reporter assays (Fig. 3a–c) and *SRC1* alternative splicing assays (Supplementary Fig. 13d), cells were grown to log phase in synthetic media containing 2% lactate as carbon source (pH 5.5). Expression by the *GAL1-10* promoter was induced for the indicated time by adding 2% galactose. The promoter was shut off by adding 2% glucose.

S. pombe strains are isogenic to JY741/746. *S. pombe* strains *Δhub1* (YHY23P) and *hub1-1* (YHY24P), and plasmids *pUR19 hub1+* and *pALSK* are gifts from H. Yashiroda[8]. Expression vectors for *S. pombe* studies were purchased from ATCC. For *S. pombe Δhub1* complementation assays, *HUB1* cDNA was cloned into *pREP81* (a vector containing thiamine-repressible *nmt81* promoter). For 5-FOA shuffle and complementation assays of *S. pombe Δsnu66*, the *SNU66* ORF with 1,000 bp of the upstream promoter and 500 bp of terminator were cloned into vectors *pUR19* (*ura4+*) or *pALSK* (*leu2+*). For complementation assays in *S. pombe*, competent cells were transformed with plasmids and incubated on selective media for 4 days at 30 °C. Fivefold serial dilutions were spotted on control and 5-FOA selection plates. The plates were incubated for 3–5 days at the temperatures indicated. Immunofluorescence in *S. pombe* was performed following a published protocol[34].

Antibodies. The following antibodies were used: anti-haemagglutinin (HA, clone F-7) and anti-Clb2 (Santa Cruz Biotechnology); anti-Myc (clone 9E10), anti-Flag (clone M2) and anti-TAP (PAP) (all from Sigma-Aldrich); sheep anti-mouse Cy3-conjugated and HRP-coupled secondary antibodies (Jackson Immuno Research); antibodies against purified recombinant Hub1 (raised in rabbit); antibodies raised against two different Snu66 peptides (Eurogentec).

Protein techniques and interaction studies. Standard procedures were followed for purification of GST- and 6×His-tagged recombinant proteins from *Escherichia coli*. Proteins were dialysed and then cleaned on size exclusion columns. Because Snu66 tends to precipitate during dialysis, the salt concentration was increased to 300 mM NaCl. Na-phosphate-based buffers were used for interaction studies, whereas HEPES-based buffer was used for proteins for structural studies. Isothermal titration calorimetry was performed to measure dissociation constants (K_d) using purified proteins and peptides. For Y2H assays, strain PJ69-7a was co-transformed with plasmids expressing the Gal4 DNA-binding domain fused to the baits, and the Gal4 activation domain fused to the preys. After 3 days of growth on selective media at 30 °C, fivefold dilutions of cultures were spotted on control plates and plates lacking the indicators histidine (-his) or adenine (-ade). The plates were further incubated for 2–5 days at 30 °C. For co-immunoprecipitation assays, 100 OD_{600} yeast cells were harvested at an $OD_{600} = 2.0$. Cells were washed once with PBS, and the pellet was frozen in liquid nitrogen. Cells were lysed either by bead beating (Retsch Instrument) or grinding in liquid nitrogen. The total yeast lysate prepared in a PBS-based buffer (1% Triton X-100, protease inhibitors by Roche), was subjected to immunoprecipitation with bead-coupled antibodies for 3 h at 4 °C. Beads were washed 4 times by rotating for 5 min at 4 °C in 1 ml lysis buffer. Bound material was eluted by a dithiothreitol-containing hydroxyurea buffer[32] followed by SDS–PAGE (4–16%) and western blot analysis. For *in vitro* GST pull-down assays, 50 μg of GST-fusion proteins bound to glutathione Sepharose beads were used to pull-down from equimolar amounts of purified proteins using the same buffer conditions as for co-immunoprecipitation experiments. Pull-down was performed for 1 h at 4 °C; subsequent washing of beads, elution of bound material and electrophoresis was similar as for the co-immunoprecipitation assay. Protein bands were stained with PageBlue (Fermentas). For HIND pull-down assays, purified Snu66-N (wild type or RR-AA mutant) was covalently coupled to Sepharose beads (GE Healthcare). Pull-down of interacting proteins from yeast lysate was performed similar to the co-immunoprecipitation experiments. For immunoprecipitations using mammalian proteins, transiently transfected 293T cells were harvested, washed in ice-cold PBS and cell pellets were lysed at 4 °C for 30 min in 5 pellet volumes of lysis buffer (50 mM HEPES pH 7.2, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 1 mM PMSF, and complete protease inhibitors (Roche)). After removal of cell debris by centrifugation (10 min, 16,000*g*, 4 °C), lysates were incubated with anti-Flag-M2 affinity gel (Sigma-Aldrich). After 3 h the affinity matrix was washed 5 times in lysis buffer and eluted in Laemmli SDS buffer.

Purification of Snu66-associated complexes. For each purification (immunoprecipitation) reaction, 200 μg anti-Snu66 antibody coupled to Dynal magnetic beads (Invitrogen) was used. Yeast lysate (~12 ml) was prepared from 3,500 OD_{600} of yeast pellet by grinding with liquid nitrogen as described[35]. Immunoprecipitation was performed at 4 °C for 3 h. Beads were washed 4 times with 12 ml lysis buffer containing 1% Triton X-100 and rotating for 5 min at 4 °C. Co-immunoprecipitation material was eluted by heating the beads at 65 °C in the presence of 2% SDS then separated on 4–12% SDS–PAGE. Protein bands were stained with PageBlue (Fermentas). Each sample lane was cut into nine gel slices, proteins were extracted, digested with trypsin, and analysed by Orbitrap mass spectrometry[36].

Splicing-specific microarrays. Splicing-specific microarrays, experimental design and data processing were described previously[20,21].

Northern blot analysis and RT–qPCR. RNA was isolated for northern blot analysis using TRIzol (Invitrogen) and for RT–qPCR using an RNeasy kit (Quiagen). ^{32}P-labelled *HUB1* and *SMT3* (SUMO)-specific probes were synthesized by random primer labelling (Ambion). Light cycler 480 was used for RT–qPCR assays (Roche). Fifteen micrograms of total RNA was used for northern blot assay whereas 500 ng of total RNA was used for reverse transcription.

Splicing reporter assays. The β-galactosidase assays were performed as described[23]. Site-directed mutagenesis was used to generate 5′-splice-site variants of the *RP51–LacZ* construct (gift from M. Rosbash). Survival on $CuSO_4$-containing solid media for *ACT1–CUP1* fusions was performed as described[24]. Yeast strain yJU75, plasmids bearing *PRP8* and *prp8-101*, and *ACT1–CUP1* reporters wild type and UuG were gifts from C. Guthrie; plasmids *prp8-R1753K*, *prp8-161* and *prp8-162*, and *ACT–CUP1* reporters A3C, BS-C and BS-G were gifts from M. Konarska.

Human cell lines, transfections and clones. HEK 293T and U2OS cells were maintained in DMEM (PAA) supplemented with 10% FCS (Biochrom AG) at 37 °C, 5% CO_2. HEK 293T cells were transfected using the calcium phosphate precipitation technique. Transfection of U2OS was performed using Lipofectamine 2000 (Invitrogen). The cDNA clone for Hs Snu66 (SART1) was purchased from Origene. The cDNA for Hs Hub1 (UBL5) was amplified by RT–PCR using total RNA from HeLa cells. Standard cloning techniques were used to generate expression constructs in *pEGFP-N1* (Clontech) or *p3×Flag-CMV-10* (Sigma-Aldrich) vectors.

Immunofluorescence microscopy. For immunofluorescence microscopy U2OS cells were seeded and transfected on glass coverslips (Roth). Cells were washed twice with PBS and fixed in 3.7% fresh paraformaldehyde (18 min, room temperature (24 °C)). After incubation, paraformaldehyde was removed by aspiration and the cells were washed three times in PBS (5 min each). Permeabilization of cells was performed using PBS-Triton X-100 (0.4% for 6 min), followed by three PBS-Tween (Tween 0.05%; PBS-T) washing steps and blocking in PBS-T/2% BSA for 1 h at room temperature. Coverslips were incubated with primary antibody for 2 h (dilution 1:200 in blocking buffer) and then washed three times in PBS-T (3 min each, room temperature). After incubation with secondary antibody (sheep anti-mouse Cy3-conjugated, Jackson Immuno Research), cover slips were mounted using DAPI-containing mounting medium (Vectashield, Vector Labs). Images were acquired on a Zeiss AxioImager Z1 microscope equipped with an AxioCam MRm Rev.3 camera. Image acquisition was carried out using AxioVision Rel. 4.7 software (Zeiss).

X-ray crystallography. Complexes of recombinant 6×His-tagged Hub1 and HIND peptides were purified on S75 Superdex column. The protein buffer used for crystallization contained 20 mM HEPES and 100 mM NaCl (pH 7.4). Crystallization of the complexes was carried out with the sitting-drop vapour diffusion method at 20 °C by mixing equal volumes of protein complex and

reservoir solution (0.2 M ammonium iodide, 20% PEG 3350, pH 6.9). Crystals of both complexes appeared in several weeks and grew to a final size of $\sim$0.3 $\times$ 0.2 $\times$ 0.1 mm. Crystals were plunged frozen after soaking for $\sim$30 s in a drop of a reservoir solution containing 30% v/v glycerol as a cryoprotectant. The crystals of the HIND-I–Hub1 complex belong to the space group $P1$, with the unit cell $a = 35.2$ Å, $b = 36.34$ Å, $c = 36.78$ Å, $\alpha = 83.44^\circ$, $\beta = 89.85^\circ$, $\gamma = 85.84^\circ$ and contained two complexes per an asymmetric unit. The HIND-II–Hub1 complex crystallized in space group $P2_12_12$, with the unit cell dimensions $a = 36.72$ Å, $b = 83.57$ Å, $c = 35.11$ Å, and contained one complex per asymmetric unit.

The data sets, up to 1.4 Å and 1.9 Å at 90 K, were collected on the MPG/GBF beamline BW6 at DESY, Hamburg, Germany using 1.05 Å wavelength. The collected data were integrated, scaled and merged by XDS and XSCALE programs[37]. The structure was determined by molecular replacement using the Molrep program from the CCP4 suite[38]. The structure of ubiquitin, taken from the PDB entry 1UBI, was used as a probe structure. Models were then refined by Refmac5 (ref. 38) and rebuilt by XtalViev/Xfit[39], followed by a subsequent Refmac5 refinement. Waters were added by Arp/warp[40]. The complete Hub1 molecule model had clear interpretable electron density, except for certain solvent exposed side chains, and those parts were omitted in the model. The structure of the Hub1–HIND-I complex has 94.9% of residues in the core, and 5.1% in allowed Ramachandran regions. For the Hub1–HIND-II complex those values are 93% and 7% respectively. Data collection and refinement statistics are summarized in Supplementary Table 2.

NMR spectroscopy. NMR spectroscopy was performed on a BRUKER AVANCE 600 MHz spectrometer equipped with a Cryo-probe. For each sample a 1H-1D spectrum with a WATERGATE-5 water suppression was measured, and for 15N labelled samples a fast HSQC spectrum[41,42] was recorded. For the full-length 2H-15N Snu66 experiment also a Trosy-HSQC was recorded[43,44].

CD spectroscopy. Synthetic peptides were solubilised in a buffer containing 20 mM NaH_2PO_4, 100 mM NaCl (pH 7.6) at 0.01–0.02 mM concentration. CD spectra were obtained on a J-715 spectropolarimeter (Jasco J715 model). All spectra were recorded by using a quartz cell with a path length of 1 mm. The parameters used for data acquisition were: response, 2 s; scanning speed, 20 nm $\min^{-1}$; bandwidth, 1.0 nm; sensitivity, 5 mdeg; and step resolution, 0.1 nm.

32. Knop, M. *et al.* Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* **15,** 963–972 (1999).
33. Janke, C. *et al.* A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21,** 947–962 (2004).
34. Hagan, I. M. & Ayscough, K. R. in *Protein Localization by Fluorescence Microscopy: A Practical Approach* (ed Allan, V. J.) 179–205 (Oxford Univ. Press, 2000).
35. Ansari, A. & Schwer, B. SLU7 and a novel activity, SSF1, act during the PRP16-dependent step of yeast pre-mRNA splicing. *EMBO J.* **14,** 4001–4009 (1995).
36. Steen, H. & Mann, M. The abc's (and xyz's) of peptide sequencing. *Nature Rev. Mol. Cell Biol.* **5,** 699–711 (2004).
37. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26,** 795–800 (1993).
38. Collaborative Computational Project 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
39. McRee, D. E. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125,** 156–165 (1999).
40. Lamzin, V. S. Automated refinement of protein models. *Acta Crystallogr. D* **49,** 129–147 (1993).
41. Liu, M. *et al.* Improved WATERGATE pulse sequences for solvent suppresion in NMR spectroscopy. *J. Magn. Reson.* **132,** 125–129 (1998).
42. Mori, S., Abeygunawardana, C., Johnson, M. O. & van Zijl, P. C. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn. Reson. B.* **108,** 94–98 (1995).
43. Zhu, G., Kong, X. M. & Sze, K. H. Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments. *J. Biomol. NMR* **13,** 77–81 (1999).
44. Pervushin, K. V., Wider, G. & Wüthrich, K. Single transition-to-single transition polarization transfer (ST2-PT) in [15N,1H]-TROSY. *J. Biomol. NMR* **12,** 345–348 (1998).

doi:10.1038/nature10137

Role of YAP/TAZ in mechanotransduction

Sirio Dupont[1]*, Leonardo Morsut[1]*, Mariaceleste Aragona[1], Elena Enzo[1], Stefano Giulitti[2], Michelangelo Cordenonsi[1], Francesca Zanconato[1], Jimmy Le Digabel[3], Mattia Forcato[4], Silvio Bicciato[4], Nicola Elvassore[2] & Stefano Piccolo[1]

Cells perceive their microenvironment not only through soluble signals but also through physical and mechanical cues, such as extracellular matrix (ECM) stiffness or confined adhesiveness. By mechanotransduction systems, cells translate these stimuli into biochemical signals controlling multiple aspects of cell behaviour, including growth, differentiation and cancer malignant progression, but how rigidity mechanosensing is ultimately linked to activity of nuclear transcription factors remains poorly understood. Here we report the identification of the *Yorkie*-homologues YAP (Yes-associated protein) and TAZ (transcriptional coactivator with PDZ-binding motif, also known as WWTR1) as nuclear relays of mechanical signals exerted by ECM rigidity and cell shape. This regulation requires Rho GTPase activity and tension of the actomyosin cytoskeleton, but is independent of the Hippo/LATS cascade. Crucially, YAP/TAZ are functionally required for differentiation of mesenchymal stem cells induced by ECM stiffness and for survival of endothelial cells regulated by cell geometry; conversely, expression of activated YAP overrules physical constraints in dictating cell behaviour. These findings identify YAP/TAZ as sensors and mediators of mechanical cues instructed by the cellular microenvironment.

Physical properties of the extracellular matrix (ECM) and mechanical forces are integral to morphogenetic processes in embryonic development, defining tissue architecture and driving specific cell differentiation programs[1]. In adulthood, tissue homeostasis remains dependent on physical cues, such that perturbations of ECM stiffness—or mutations affecting its perception—are causal to pathological conditions of multiple organs, contributes to ageing and cancer malignant progression[2].

Mechanotransduction enables cells to sense and adapt to external forces and physical constraints[3,4]; these mechanoresponses involve the rapid remodelling of the cytoskeleton, but also require the activation of specific genetic programs. In particular, variations of ECM stiffness or changes in cell shape caused by confining the cell's adhesive area have a profound impact on cell behaviour across several cell types, such as mesenchymal stem cells[5,6], muscle stem cells[7] and endothelial cells[8]. The nuclear factors mediating the biological response to these physical inputs remain incompletely understood.

ECM stiffness regulates YAP/TAZ activity

To gain insight into these issues, we asked if physical/mechanical stimuli conveyed by ECM stiffness actually signal through known signalling pathways. For this, we performed a bioinformatic analysis on genes differentially expressed in mammary epithelial cells (MEC) grown on ECM of high versus low stiffness[9]. Specifically, we searched for statistical associations between genes regulated by stiffness and gene signatures denoting the activation of specific signalling pathways (Supplementary Fig. 2, Supplementary Table 1 and Methods). We included signatures of MAL/SRF and NF-κB as these factors translocate in the nucleus in response to changes in F-actin polymerization and cell stretching[10]. Strikingly, only signatures revealing activation of YAP/TAZ transcriptional regulators emerged as significantly overrepresented in the set of genes regulated by high stiffness (Supplementary Fig. 2).

To test if YAP and TAZ activity is regulated by ECM stiffness, we monitored YAP/TAZ transcriptional activity in human MEC grown on fibronectin-coated acrylamide hydrogels of varying stiffness (elastic modulus ranging from 0.7 to 40 kPa, matching the physiological elasticities of natural tissues[6]). For this, we assayed by real-time PCR two of the best YAP/TAZ regulated genes from our signature, *CTGF* and *ANKRD1*. The activity of YAP/TAZ in cells grown on stiff hydrogels (15–40 kPa) was comparable to that of cells grown on plastics, whereas growing cells on soft matrices (in the range of 0.7–1 kPa) inhibited YAP/TAZ activity to levels comparable to short interfering RNA (siRNA)-mediated YAP/TAZ depletion (Fig. 1a and data not shown). We confirmed this finding in other cellular systems, such as MDA-MB-231 and HeLa cells, where we used a synthetic YAP/TAZ-responsive luciferase reporter (4×GTIIC-lux) as direct read-out of their activity (Fig. 1a and Supplementary Fig. 4).

Next, we assayed endogenous YAP/TAZ subcellular localization; indeed, their cytoplasmic relocalization has been extensively used as primary read-out of their inhibition by the Hippo pathway or by cell–cell contact (Supplementary Fig. 5 and ref. 11). By immunofluorescence on MEC and human mesenchymal stem cells (MSC, an established non-epithelial cellular model for mechanoresponses[5,6]), YAP/TAZ were clearly nuclear on hard substrates but became predominantly cytoplasmic on softer substrates (Fig. 1b and Supplementary Figs 6 and 7). Collectively, these data indicate that YAP/TAZ activity and subcellular localization are regulated by ECM stiffness.

YAP/TAZ are regulated by cell geometry

It is recognized that changes in ECM stiffness impose different degrees of cell spreading[6,12]. We thus asked whether cell spreading is sufficient to regulate YAP/TAZ. To this end, we used micropatterned fibronectin 'islands' of defined size, on which cells can spread to different degrees depending on the available adhesive area[8]. On these micropatterns, the localization of YAP/TAZ changed from predominantly nuclear in spread MSCs, to predominantly cytoplasmic in cells on smaller islands (Fig. 1c). Of note, the use of single-cell adhesive islands rules out the possibility that cell–cell contacts could be involved in YAP/TAZ relocalization. We confirmed these results using human lung microvascular endothelial cells (HMVEC, Fig. 1d), that are well known to regulate their growth according to cell shape[8].

[1]Department of Histology, Microbiology and Medical Biotechnologies, University of Padua School of Medicine, viale Colombo 3, 35131 Padua, Italy. [2]Department of Chemical Engineering (DIPIC), University of Padua, via Marzolo 9, 35131 Padua, Italy. [3]Laboratoire Matière et Systèmes Complexes (MSC), Université Paris Diderot and CNRS UMR 7057, 10 rue A. Dumont et L. Duquet, 75205 Paris, France. [4]Center for Genome Research, Department of Biomedical Sciences, University of Modena and Reggio Emilia, via G. Campi 287, 41100 Modena, Italy.
*These authors contributed equally to this work.

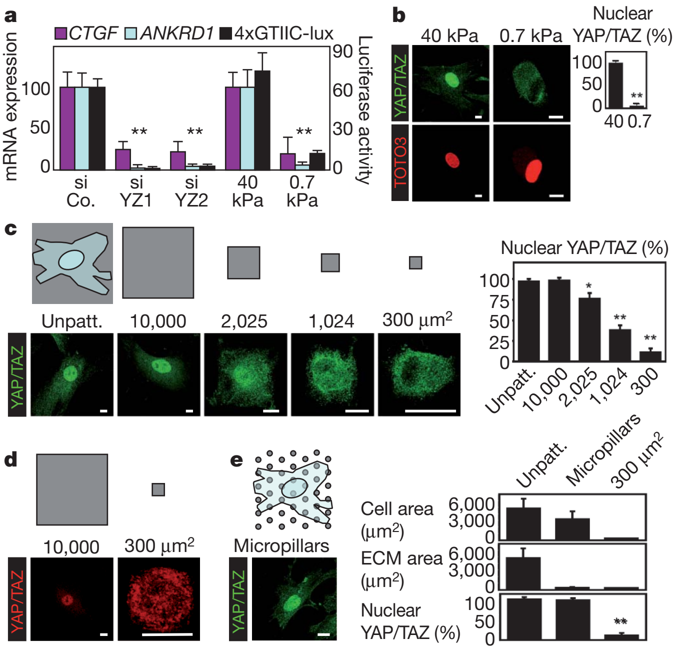


Figure 1 | YAP/TAZ are regulated by ECM stiffness and cell shape a, Real-time PCR analysis in MCF10A cells (*CTGF* and *ANKRD1*, coloured bars) and luciferase reporter assay in MDA-MB-231 cells (4×GTIIC-lux, black bars) to measure YAP/TAZ transcriptional activity. Cells were transfected with the indicated siRNAs (siCo., control siRNA; siYZ1 and siYZ2, two YAP/TAZ siRNAs; see Supplementary Fig. 3) and grown on plastic, or plated on stiff (elastic modulus of 40 kPa) and soft (0.7 kPa) fibronectin-coated hydrogels. Data are normalized to lane 1. $n = 4$. **b**, Confocal immunofluorescence images of YAP/TAZ and nuclei (TOTO3) in human mesenchymal stem cells (MSC) plated on hydrogels. Scale bars, 15 μm. Graphs indicate the percentage of cells with nuclear YAP/TAZ. ($n = 3$). **c**, On top: grey patterns show the relative size of microprinted fibronectin islands on which cells were plated. Outline of a cell is shown superimposed to the leftmost unpatterned area (Unpatt.). Below: confocal immunofluorescence images of MSC plated on fibronectin islands of decreasing sizes (μm^2). Scale bars, 15 μm. Graph provides quantifications. ($n = 8$). See also Supplementary Fig. 8. **d**, Confocal immunofluorescence images of YAP/TAZ in HMVEC plated as in **c**. Scale bars, 15 μm. See also Supplementary Fig. 9. **e**, On top: grey dots exemplify the distribution of fibronectin on micropillar arrays, shown superimposed with the outline of a cell. Below: representative immunofluorescence of YAP/TAZ in MSC plated on rigid micropillars. Scale bars, 15 μm. Graphs, quantification of the projected cell area, total ECM contact area, and nuclear YAP/TAZ in MSC plated on unpatterned fibronectin, micropillars and 300 μm^2 islands. ($n = 4$). All error bars are s.d. (*$P < 0.05$; **$P < 0.01$; Student's *t*-test is used throughout). Experiments were repeated *n* times with duplicate biological replicates.

Cells seeded on stiff hydrogels or large islands show increased cell spreading but, at the same time, experience a broader cell–ECM contact area. To test whether YAP/TAZ are regulated by cell spreading irrespectively of the total amount of ECM, we visualized YAP/TAZ localization in MSC grown on the tip of closely arrayed fibronectin-coated micropillars[12]: on these arrays, cells stretch from one micropillar to another, and assume a projected cell area comparable to cells plated on big islands (3,200 μm^2 on average, Fig. 1e); however, in these conditions, the actual area available for cell–ECM interaction is only about 10% of their projected area (300 μm^2 on average, corresponding to the smallest islands used in Fig. 1c). YAP/TAZ remained nuclear on micropillars (Fig. 1e), indicating that YAP/TAZ are primarily regulated by cell spreading imposed by the ECM.

YAP/TAZ sense cytoskeletal tension

We then considered that cell spreading entails activation of the small GTPase Rho that, in turn, regulates the formation of actin bundles, stress fibres and tensile actomyosin structures[2,3]. Indeed, cells on stiff ECM or big islands had more prominent stress fibres compared to those plated on soft ECM or small islands (Supplementary Figs 9 and 10). As shown in Fig. 2a, we found that Rho and the actin cytoskeleton are required to maintain nuclear YAP/TAZ in MSC. As a control, inhibition of Rac1-GEFs (guanine nucleotide exchange factors), or disruption of microtubules, did not alter YAP/TAZ localization (Fig. 2a). Similar results were obtained also in HMVEC and MEC (not shown). Crucially, inhibition of Rho and of the actin cytoskeleton also inhibited YAP/TAZ transcriptional activity, as assayed by expression of endogenous target genes (Fig. 2b) and by luciferase reporter assays (Fig. 2c). Conversely, triggering F-actin polymerization and stress fibres formation by overexpression of activated diaphanous protein (*DIAPH1*) promoted YAP/TAZ activity (Supplementary Fig. 12).

We then asked whether YAP/TAZ are regulated by the ratio of monomeric/filamentous actin, as others observed for MAL/SRF[13]. To increase monomeric G-actin, we overexpressed the R62D mutant actin[13], but this was insufficient to inhibit YAP/TAZ (Fig. 2c). Moreover, increasing the amount of F-actin either by overexpressing the F-actin stabilizing V159N actin mutant or by serum stimulation[13] had no effect on YAP/TAZ activity (Supplementary Fig. 13) or nuclear localization (data not shown). As a control, in the same experimental set-up, MAL/SRF activity was instead clearly modulated

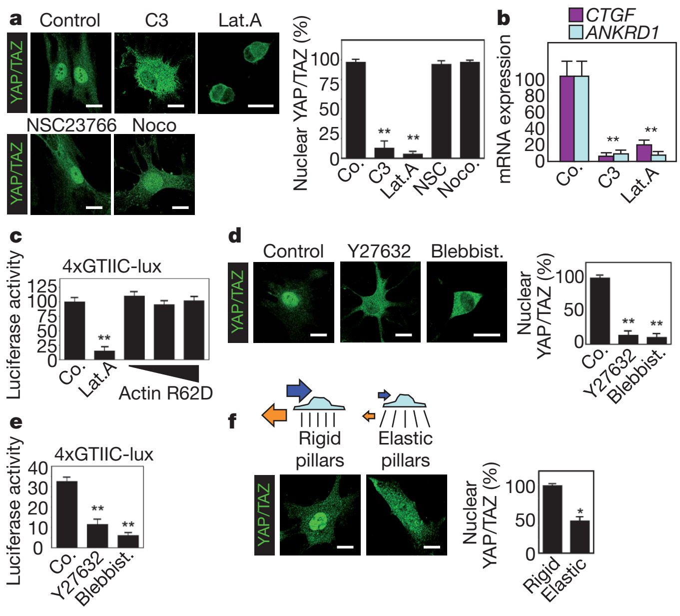


Figure 2 | YAP/TAZ activity requires Rho and tension of the actin cytoskeleton a, Confocal immunofluorescence images of YAP/TAZ in MSC treated with the Rho inhibitor C3 (3 $\mu g\, ml^{-1}$), the F-actin inhibitor latrunculin A (Lat.A, 0.5 μM), the Rac1-GEFs inhibitor NSC23766 (100 μM) or the microtubule inhibitor nocodazole (Noco., 30 μM). Scale bars, 15 μm. Graph provides quantifications ($n = 10$). See also Supplementary Fig. 11. **b**, Real-time PCR of MCF10A treated with cytoskeletal inhibitors as in **a**. Data are normalized to untreated cells (Co.) ($n = 4$). **c**, Luciferase assay for YAP/TAZ activity in HeLa cells transfected with the indicated expression plasmids (Co. is empty vector, actin R62D encodes for a mutant unable to polymerize into F-actin) and treated with latrunculin A ($n = 4$). Similar effects were observed in MDA-MB-231 (not shown). **d**, Confocal immunofluorescence images of MSC treated with the ROCK inhibitor Y27632 (50 μM), or the non-muscle myosin inhibitor blebbistatin (Blebbist., 50 μM). Scale bars, 15 μm. Graph provides quantifications ($n = 9$). See also Supplementary Fig. 15. **e**, Luciferase activity of the YAP/TAZ reporter in HeLa treated as in **d**. ($n = 4$). **f**, Confocal immunofluorescence images of MSC plated on arrays of micropillars of different rigidities. On rigid micropillars (black lines) cells develop cytoskeletal tension (blue arrow) by pulling against the ECM (orange arrow); cells bend elastic micropillars and develop reduced tension exemplified by reduced size of the arrows. Scale bars, 15 μm. Graph provides quantifications ($n = 2$). See also Supplementary Fig. 19. All error bars are s.d. (*$P < 0.05$; **$P < 0.01$). Experiments were repeated *n* times with duplicate biological replicates.

(Supplementary Fig. 14). Taken altogether, these data indicate that Rho and stress fibres, but not F-actin polymerization *per se*, are required for YAP/TAZ activity.

Cells respond to the rigidity of the ECM by adjusting the tension and organization of their stress fibres, such that cell spreading is accompanied by increased pulling forces against the ECM[3,6,12]. By inhibition of ROCK and non-muscle myosin[4,6], we found that cytoskeletal tension is required for YAP/TAZ nuclear localization (Fig. 2d) and activity (Fig. 2e and Supplementary Fig. 16). Of note, YAP/TAZ exclusion caused by these inhibitions is an early event (occurring within 2 h) that can be uncoupled from destabilization of stress fibres (see Supplementary Fig. 17). By comparison, the activity of MAL/SRF was only marginally affected by the same treatments (Supplementary Fig. 18). To address more directly the relevance of cell-generated mechanical force without using small-molecule inhibitors and irrespectively of the surface properties of the hydrogels, we compared rigid versus highly elastic micropillars[12]; on the elastic substrate, cytoplasmic localization of YAP/TAZ was clearly increased (Fig. 2f). Collectively, the data indicate that YAP/TAZ respond to cytoskeletal tension.

We also tested if inhibition of YAP/TAZ occurs by entrapping YAP/TAZ in the cytoplasm or by promoting their nuclear exclusion. As shown in Supplementary Fig. 20a, blockade of nuclear export with leptomycin B rescued nuclear localization of YAP/TAZ in MSC treated with cytoskeletal inhibitors, indicating that YAP/TAZ keeps shuttling between cytoplasm and nucleus irrespectively of cell tension, and that the presence of a tense cytoskeleton promotes their nuclear retention. Moreover, YAP/TAZ relocalization was rapid (occurring in as little as 30 min with latrunculin A), reversible after small-molecule washout (Supplementary Fig. 20b), and insensitive to inhibition of protein synthesis with cycloheximide (data not shown), suggesting a direct biochemical mechanism.

Mechanical cues act independently from Hippo

YAP and TAZ are the nuclear transducers of the Hippo pathway[14]. In several organisms and cellular set-ups, activation of the Hippo pathway leads to YAP/TAZ phosphorylation on specific serine residues; in turn, these phosphorylations inhibit YAP/TAZ activity through multiple mechanisms, including proteasomal degradation[14]. Intriguingly, similar to Hippo activation by cell–cell contacts (Fig. 3a), TAZ protein was also degraded by growing MEC on soft matrices (Fig. 3b) or by treatment with inhibitors of Rho, F-actin and actomyosin tension (Fig. 3c and Supplementary Fig. 21). Similar results were obtained with MSC and HMVEC (Supplementary Fig. 22 and data not shown).

Is then the Hippo cascade responsible for YAP/TAZ inhibition by mechanical cues? Several evidences indicate this is not the case. First, we noted that phosphorylation of YAP on serine 127, a key target of the LATS kinase downstream of the Hippo pathway[15], was not increased upon treatment of MEC and MSC with cytoskeletal inhibitors (Fig. 3c and Supplementary Fig. 22), at difference with its regulation by high confluence (see Fig. 3a). Second, depletion of LATS1 and LATS2 (see Fig. 3f and Supplementary Fig. 23 for positive controls) had marginal effect on YAP/TAZ inactivation by mechanical cues, as judged by (1) YAP/TAZ nuclear exit induced by micropatterns or cytoskeletal inhibition in MEC, MSC or HMVEC (Fig. 3d and Supplementary Figs 24, 25 and data not shown); (2) TAZ degradation (Fig. 3e); (3) endogenous target gene expression in cells plated on soft hydrogels (Fig. 3f) or treated with latrunculin A (Supplementary Fig. 26). Finally, we compared wild-type or LATS-insensitive 4SA[16] TAZ in MDA-MB-231 depleted of endogenous YAP/TAZ and reconstituted at near-to-endogenous YAP/TAZ activity levels with siRNA-insensitive mouse TAZ vectors. As shown in Fig. 3g, both wild-type and 4SA TAZ remain sensitive to mechanical cues. Further supporting these results, we found that MDA-MB-231 cells are homozygous mutant for NF2 (also known as merlin, Supplementary Fig. 27), an essential component of the Hippo cascade[14]. Collectively, these data

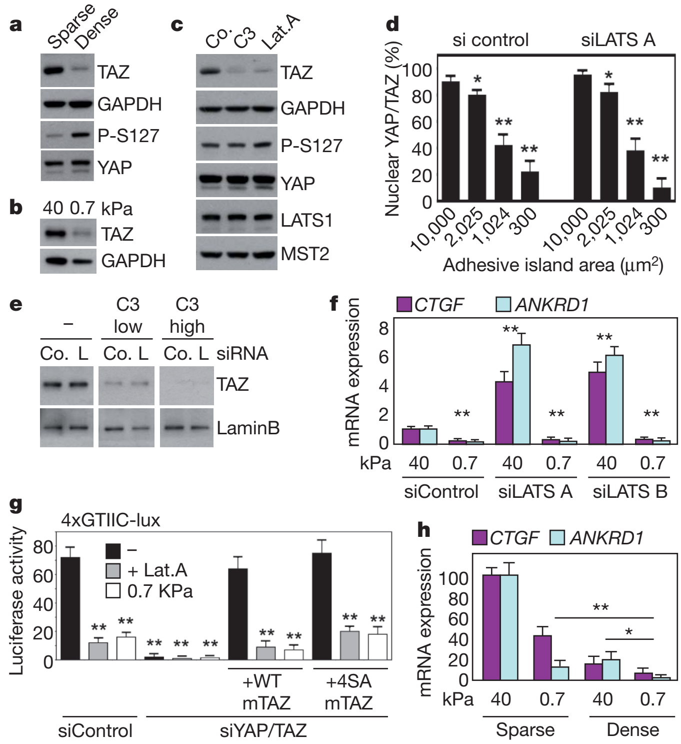


Figure 3 | ECM stiffness and cell spreading regulate YAP/TAZ independently of the Hippo pathway **a**–**c**, Immunoblotting for the indicated proteins in MCF10A under the following conditions: **a**, plating on plastic at low (sparse) or high (dense) confluence; **b**, plating on stiff (40 kPa) or soft (0.7 kPa) hydrogels; **c**, untreated (Co.) or treated with C3 and latrunculin A (Lat.A). P-S127 is phospho-YAP. **d**, Quantification of nuclear YAP/TAZ in MSC transfected with control or LATS1/2 siRNA A (siLATS A) and plated on microprinted islands of different size ($n = 4$). Similar results were obtained with HMVEC (not shown). **e**, Immunoblotting from MSC cells transfected with the indicated siRNAs (Co., control siRNA; L, LATS1/2 siRNA A), plated on plastic and treated with C3 (0.5 or 3 μg ml^{-1}). Similar results were obtained by using blebbistatin or latrunculin A, or by treating HMVEC and MCF10A (not shown). **f**, Real-time PCR analysis of MCF10A transfected with the indicated siRNAs and cultured on hydrogels. Data are normalized to the first lane ($n = 3$). **g**, Luciferase assay in MDA-MB-231 transfected as indicated and treated with latrunculin A (Lat.A) or replated on soft hydrogels. ($n = 8$). **h**, RT–PCR of MCF10A grown under sparse or confluent (dense) conditions on the indicated hydrogels. Data are normalized to the first lane ($n = 2$). All error bars are s.d. (*$P < 0.05$; **$P < 0.01$). Experiments were repeated n times with duplicate biological replicates.

indicate that LATS phosphorylation downstream of the Hippo cascade is not the primary mediator of mechanical/physical cues in regulating YAP/TAZ activity.

We then asked if mechanical cues regulate YAP/TAZ not only in isolated cells, but also in confluent monolayers, when cells reorganize their shape and structure and engage in cell–cell contacts, leading to activation of Hippo/LATS signalling[11]. We first explored the effects of cell confluence in a simplified experimental set-up, namely in MCF10A cells rendered insensitive to Hippo activation by depletion of LATS1/2; in these conditions, Rho and the cytoskeleton remain relevant inputs to support TAZ stability (Supplementary Fig. 28). Moreover, in parental MCF10A, plating cells at high confluence cooperate with soft hydrogels in inhibiting YAP/TAZ activity (Fig. 3h). Thus, mechanical cues and Hippo signalling represent two parallel inputs converging on YAP/TAZ regulation.

YAP/TAZ mediate cellular mechanoresponses

Data presented so far indicate YAP and TAZ as molecular 'readers' of ECM elasticity and cell geometry, but are YAP/TAZ relevant to mediate the biological responses to these mechanical inputs? An appropriate

cellular model to address this question are MSC, that can differentiate into osteoblasts when cultured on stiff ECM, mimicking the natural bone environment, whereas on soft ECM—or small islands—they differentiate into other lineages, such as adipocytes[5,6]. A similar case applies to endothelial cells, that respond differently to the same soluble growth factor by proliferating, differentiating or involuting according to the degree of cell spreading against the surrounding ECM[8]. We proposed that cell fates induced by stiff ECM and large islands (that is, where YAP/TAZ are active) should require YAP/TAZ function and, conversely, cell fates associated to soft ECM and small islands (where YAP/TAZ are inhibited) should require their inactivation. In line with this hypothesis, osteogenic differentiation induced in MSC on stiff ECM was inhibited upon depletion of YAP and TAZ, and a similar inhibition was achieved either by culturing cells on soft ECM or by incubation with C3 (Fig. 4a, b). We also monitored adipogenic differentiation, a fate normally not allowed on stiff ECM; strikingly, YAP/TAZ knockdown enabled adipogenic differentiation on stiff substrates, thus mimicking a soft environment (Fig. 4c and Supplementary Fig. 30). In the case of HMVEC, cells plated on small islands undergo apoptosis, whereas cells on bigger islands proliferate, as assayed by TdT-mediated dUTP nick end labelling (TUNEL) staining and 5-bromodeoxyuridine (BrdU) incorporation, respectively[8]. Upon YAP/TAZ depletion, cells on bigger islands behaved as if they were on small islands; this is overlapping with the biological effects of Rho inhibition (Fig. 4d). In line with the Hippo independency of this regulation, knockdown of LATS1/2 was not sufficient to rescue osteogenesis upon C3 treatment, or endothelial cell proliferation on small islands (Supplementary Fig. 32). Collectively, these data suggest that YAP/TAZ are required for cell differentiation triggered by changes in ECM stiffness and for geometric control of cell survival.

We next tested if the sole YAP/TAZ activity can re-direct the biological responses elicited by soft/confined ECM. Overexpression of activated 5SA-YAP (ref. 14) with lentiviral infection (to at least tenfold the endogenous levels, data not shown) remarkably overruled the geometric control over proliferation and apoptosis in HMVEC (Fig. 4e), and rescued osteogenic differentiation of MSC treated with C3 or plated on soft ECM (Fig. 4f, g). Thus, cells on soft matrices or on small adhesive areas can be 'tricked' to behave as if they were adhering on harder/larger substrates by sustaining YAP/TAZ function.

Discussion

In summary, our findings indicate a fundamental role of the transcriptional regulators YAP and TAZ as downstream elements in how cells perceive their physical microenvironment (Supplementary Fig. 1). Our data define an unprecedented modality of YAP/TAZ regulation, that acts in parallel to the NF2/Hippo/LATS pathway, and instead requires Rho activity and the actomyosin cytoskeleton. Interestingly, this recapitulates aspects of MAL/SRF regulation[13], but also entails profound differences: YAP/TAZ activity requires stress fibres and cytoskeletal tension induced by ECM stiffness and cell spreading, but is not directly regulated by G-actin levels. The detailed biochemical mechanisms by which cytoskeletal tension regulates YAP/TAZ await further characterization, but it is tempting to speculate that stress-fibres inhibit an unidentified YAP/TAZ-antagonist. Functionally, we showed in different cellular models that cells read ECM elasticity, cell shape and cytoskeletal forces as levels of YAP/TAZ activity, such that experimental manipulations of YAP/TAZ levels can dictate cell behaviour, overruling mechanical inputs. This identifies a new widespread transcriptional mechanism by which the mechanical properties of the ECM and cell geometry instruct cell behaviour. This may now shed light on how physical forces shape tissue morphogenesis and homeostasis, for example in tissues undergoing constant remodelling upon variation of their mechanical environment; indeed, alterations of YAP/TAZ signalling have been genetically linked in animal models to the emergence of cystic kidney, pulmonary emphysema, heart and vascular defects[17–20]. In cancer, changes in the ECM composition and

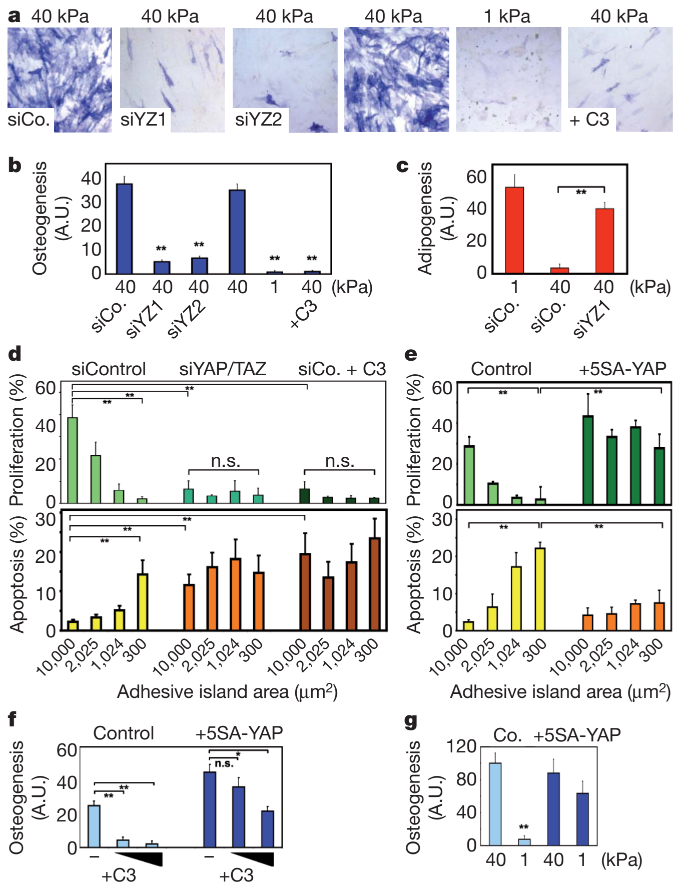


Figure 4 | YAP/TAZ are required mediators of the biological effects controlled by ECM elasticity and cell geometry a–c, MSC were transfected with the indicated siRNA (control, siCo.; YAP/TAZ, siYZ1 and siYZ2), plated on stiff (40 kPa) or soft (1 kPa) substrates and induced to differentiate into osteoblasts (**a**, **b**) or adipocytes (**c**). C3 (0.5 μg ml^{-1}) was added and renewed with differentiation medium. **a**, Representative alkaline phosphatase stainings and **b**, quantifications of osteogenic differentiation ($n = 4$). **c**, Quantification of adipogenesis based on oil-red stainings ($n = 2$) (A.U., arbitrary units, see methods). See Supplementary Figs 29 and 30 for controls. These results are consistent with ref. 23. **d**, Proliferation (BrdU, upper panel) and apoptosis (TUNEL, lower panel) of HMVEC plated on adhesive islands of different size; where indicated, cells were treated overnight with C3 (2.5 μg ml^{-1}), or transfected with the indicated siRNAs ($n = 5$). Similar results were obtained with siYZ2 (not shown). Representative stainings in Supplementary Fig. 31. **e**, Proliferation (upper panel) and apoptosis (lower panel) of control and 5SA-YAP-expressing HMVEC, plated on adhesive islands. **f**, **g**, Quantifications of osteogenesis in MSC transduced with 5SA-YAP, and treated with C3 (50 and 150 ng ml^{-1}) ($n = 3$) (**f**) or plated on hydrogels ($n = 2$) (**g**). Representative stainings in Supplementary Fig. 33. All error bars are s.d. ($*P < 0.05$; $**P < 0.01$; n.s., not significant). Experiments were repeated n times with duplicate biological replicates.

mechanical properties is the focus of intense interest, as these have been correlated with progression and build-up of the metastatic niche[2]; in light of their powerful oncogenic activities[14], YAP/TAZ might serve as executers of these malignant programs.

Genetically, YAP and TAZ have been linked to a universal system that controls organ size[14]. The current view implicates Hippo signalling as the sole determinant of YAP/TAZ regulation in tissues. However, our results suggest physical/mechanical inputs as alternative determinants of YAP/TAZ activity. Supporting this view, it has been observed that growth of epithelial tissues entails the build-up of mechanical stresses at tissue boundaries[21], and theoretical work proposed that these serve as positive feedback to homogenize cell growth, compensating for uneven activity of soluble growth factors[22]. It is tempting to speculate that proliferative tissue homeostasis may be

achieved by a combination of growth factor signalling and localized control of YAP/TAZ activation by cell–cell contacts and mechanical cues dictated by tissue architecture.

METHODS SUMMARY

MSC and HMVEC-L cells, their growth and differentiation media, were from Lonza. Micropatterned slides were from Cytoo SA. Micropost arrays and acrylamide hydrogels were synthesized according to standard procedures. Drug treatments were performed in 8-well Lab-Tek chamber slides (Nunc). Transfections were carried out with TransitLT1 (MirusBio) for plasmids, with Lipofectamine RNAiMax (Invitrogen) for siRNA (sequences in Supplementary Table 2). Anti-YAP/TAZ is sc101199 (SantaCruz). Other stainings were DeadEnd (Promega) for TUNEL, kit number 1 (Roche) for BrdU, kit number 85L2 (Sigma) for alkaline phosphatase, Oil-red (Sigma) for lipid vacuoles. Real-time PCR was performed on dT-primed cDNA with a RG3000 Corbett Research cycler (primers in Supplementary Table 3).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 November 2010; accepted 19 April 2011.

1. Mammoto, T. & Ingber, D. E. Mechanical control of tissue and organ development. *Development* **137,** 1407–1420 (2010).
2. Jaalouk, D. E. & Lammerding, J. Mechanotransduction gone awry. *Nature Rev. Mol. Cell Biol.* **10,** 63–73 (2009).
3. Schwartz, M. A. Integrins and extracellular matrix in mechanotransduction. *Cold Spring Harb. Perspect. Biol.* **2,** a005066 (2010).
4. Vogel, V. & Sheetz, M. Local force and geometry sensing regulate cell functions. *Nature Rev. Mol. Cell Biol.* **7,** 265–275 (2006).
5. McBeath, R., Pirone, D. M., Nelson, C. M., Bhadriraju, K. & Chen, C. S. Cell shape, cytoskeletal tension, and RhoA regulate stem cell lineage commitment. *Dev. Cell* **6,** 483–495 (2004).
6. Engler, A. J., Sen, S., Sweeney, H. L. & Discher, D. E. Matrix elasticity directs stem cell lineage specification. *Cell* **126,** 677–689 (2006).
7. Gilbert, P. M. *et al.* Substrate elasticity regulates skeletal muscle stem cell self-renewal in culture. *Science* **329,** 1078–1081 (2010).
8. Chen, C. S., Mrksich, M., Huang, S., Whitesides, G. M. & Ingber, D. E. Geometric control of cell life and death. *Science* **276,** 1425–1428 (1997).
9. Provenzano, P. P., Inman, D. R., Eliceiri, K. W. & Keely, P. J. Matrix density-induced mechanoregulation of breast cell phenotype, signaling and gene expression through a FAK-ERK linkage. *Oncogene* **28,** 4326–4343 (2009).
10. Olson, E. N. & Nordheim, A. Linking actin dynamics and gene transcription to drive cellular motile functions. *Nature Rev. Mol. Cell Biol.* **11,** 353–365 (2010).
11. Zhao, B. *et al.* Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes Dev.* **21,** 2747–2761 (2007).
12. Fu, J. *et al.* Mechanical regulation of cell function with geometrically modulated elastomeric substrates. *Nature Methods* **7,** 733–736 (2010).
13. Miralles, F., Posern, G., Zaromytidou, A. I. & Treisman, R. Actin dynamics control SRF activity by regulation of its coactivator MAL. *Cell* **113,** 329–342 (2003).
14. Pan, D. The hippo signaling pathway in development and cancer. *Dev. Cell* **19,** 491–505 (2010).
15. Oka, T., Mazack, V. & Sudol, M. Mst2 and Lats kinases regulate apoptotic function of Yes kinase-associated protein (YAP). *J. Biol. Chem.* **283,** 27534–27546 (2008).
16. Lei, Q. Y. *et al.* TAZ promotes cell proliferation and epithelial-mesenchymal transition and is inhibited by the hippo pathway. *Mol. Cell. Biol.* **28,** 2426–2436 (2008).
17. Chen, Z., Friedrich, G. A. & Soriano, P. Transcriptional enhancer factor 1 disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. *Genes Dev.* **8,** 2293–2301 (1994).
18. Makita, R. *et al.* Multiple renal cysts, urinary concentration defects, and pulmonary emphysematous changes in mice lacking TAZ. *Am. J. Physiol. Renal Physiol.* **294,** F542–F553 (2008).
19. Morin-Kensicki, E. M. *et al.* Defects in yolk sac vasculogenesis, chorioallantoic fusion, and embryonic axis elongation in mice with targeted disruption of *Yap65*. *Mol. Cell. Biol.* **26,** 77–87 (2006).
20. Skouloudaki, K. *et al.* Scribble participates in Hippo signaling and is required for normal zebrafish pronephros development. *Proc. Natl Acad. Sci. USA* **106,** 8579–8584 (2009).
21. Nienhaus, U., Aegerter-Wilmsen, T. & Aegerter, C. M. Determination of mechanical stress distribution in *Drosophila* wing discs using photoelasticity. *Mech. Dev.* **126,** 942–949 (2009).
22. Schwank, G. & Basler, K. Regulation of organ growth by morphogen gradients. *Cold Spring Harb. Perspect. Biol.* **2,** a001669 (2010).
23. Hong, J. H. *et al.* TAZ, a transcriptional modulator of mesenchymal stem cell differentiation. *Science* **309,** 1074–1078 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Scita for advice and gift of reagents; X. Yang for 5SA-YAP1 plasmid; I. Farrance for 4×GTIIC-lux plasmid; H. Miyoshi for pCSII-EF-MCS vector; L. Naldini for pMD2-VSVG vector; R. Treisman for ΔN1+C mDIA, R26D and V159N Actin plasmids; G. Posern for SRF-lux reporter; mouse TAZ and psPAX2 were Addgene plasmid 19025 and 12260. This work was supported by: Telethon and Progetti di Eccellenza CARIPARO grants to N.E.; AIRC (Italian Association for Cancer Research) PI and AIRC Special Program Molecular Clinical Oncology "5 per mille", University of Padua Strategic grant, IIT Excellence grant and Telethon to S.P.; AIRC PI and MIUR (Italian Minister of University) grants to S.D.

Author Contributions S.D., L.M. and S.P. designed research; L.M., S.D., M.A., E.E. and F.Z. performed experiments; M.C., S.B. and M.F. performed bioinformatics analysis; N.E. and S.G. prepared hydrogels; J.LeD. prepared micropost arrays; S.D. and S.P. coordinated the project; S.D. and S.P. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.D. (dupont@bio.unipd.it) and S.P. (piccolo@bio.unipd.it).

METHODS

Reagents, microfabrications and plasmids. Cell-permeable C3 transferase (Cytoskeleton Inc.) was used in serum-free conditions for MCF10A and MSC, in complete medium for HMVEC. Y27632, blebbistatin, nocodazole were from Sigma. Latrunculin A was from Santa Cruz. NSC23766 was from Tocris. Micropatterned glass slides were purchased from Cytoo SA; on every slide, square islands of different sizes were arrayed in quadrants, leaving 70 μm of non-adhesive glass between each island; a control area evenly coated with fibronectin was also included to let cells attach without geometric constraints. Fibronectin-coated hydrogels were synthesized according to ref. 24. Micropost arrays were prepared according to ref. 25, with microposts of 1 μm in diameter and 3 μm of centre-to-centre distance; elasticity of the micropillars was changed by modulating the amount of cross-linker (10% in the stiff micropillars, 5% in the elastic ones) and their length, as in ref. 12, to obtain nominal spring constants of $>$10.9 nN μm^{-1} for rigid micropillars, and 1.39 nN μm^{-1} for the elastic ones. 5SA-YAP1 was subcloned into pCSII-EF-MCS to produce lentiviral particles. 4SA-mouse TAZ cDNA was synthesized ad hoc (GeneScript).

Cell lines, transfections and treatments. Mouse NMuMG cells were grown in DMEM 10% FCS. Human MCF10A cells were grown in DMEM/F12 with 5% horse serum freshly supplemented with insulin, epidermal growth factor, hydrocortisone and cholera toxin. Human MDA-MB-231 cells were grown in DMEM/F12 with 10% FBS. Bone marrow-derived MSC and HMVEC-L were purchased from Lonza and grown according to the manufacturer's instructions. siRNA transfections were done with Lipofectamine RNAi-MAX (Invitrogen) in antibiotics-free medium according to manufacturer's instructions. Sequences of siRNAs is provided in Supplementary Table 2. DNA transfections were done with TransitLT1 (Mirus Bio). Lentiviral particles were prepared by transiently transfecting HEK293T cells with lentiviral vectors together with packaging vectors (pMD2-VSVG and psPAX2). Luciferase assays with the established YAP/TAZ-responsive reporter 4×GTIIC-lux[26] were as in ref. [27], and displayed as arbitrary units.

For hydrogels, 5,000–10,000 cells per cm^2 were seeded in drop; after attachment, the wells containing the hydrogels were filled with appropriate medium. MSC and mammary cells were plated in growth medium and harvested for immunofluorescence after 24 h; for luciferase and gene expression after 48 h. For bone differentiation assays, growth medium was changed with osteogenic differentiation medium 24 h after seeding, and renewed every 2 days for a total of 8 days of differentiation. Bone differentiation was assayed by alkaline phosphatase staining (Sigma 85L2) and quantified with ImageJ software as follows: for each sample, at least five low magnification (×20) pictures were taken, and the alkaline-phosphatase-positive area was determined with ImageJ as the number of blue pixels across the picture; this value was then normalized to the number of cells (Hoechst/nuclei) for each picture (arbitrary units). For adipogenic differentiation, growth medium was replaced with adipogenic induction medium 24 h after seeding; cells were then subjected to cycles of 3 days of adipogenic induction and 1 day of adipogenic maintenance until harvesting at day 7 of differentiation. Adipogenic differentiation was assayed by Oil Red staining (Sigma) and quantified as the Oil Red-positive area normalized to the number of cells (Hoechst-positive nuclei) in a manner similar to that described for bone differentiation.

For micropatterns and micropost arrays, 40,000 HMVEC or MSC cells were plated in 35-mm dishes in growth medium. For immunofluorescence, cells were fixed 24 h after plating. For HMVEC proliferation and apoptosis assays, cells were fixed 24 h after plating (including 1 h incubation with BrdU in the case of proliferation assays) and processed according to TUNEL or BrdU detection kits (Promega DeadEnd and Roche Kit number 1, respectively). The projected cell area of cells on fibronectin-coated glass slides and on microposts was determined with imageJ based on immunofluorescence pictures of cells stained with anti-YAP/TAZ; the area of ECM contacted by cells was estimated by calculating that microposts (diameter 1 μm) arrayed in equilateral triangles (centre-to-centre 3 μm) approximate 10% of the total surface covered by cells (projected cell area).

For drug treatments and immunofluorescence, 10,000 cells per cm^2 were plated onto 8-well glass Lab-Tek chamberslides (Nunc) precoated for 1 h at 37 °C with 20 μg ml^{-1} bovine fibronectin (Sigma) in 1× PBS. Unless indicated otherwise, drug concentrations are indicated in the legend to Fig. 2a and d, and treatments lasted 4 h for immunofluorescence, 6 h for western blotting, and overnight for luciferase and gene expression assays. For serum stimulations, cells were incubated overnight without serum and then stimulated for 6 h with 20% serum; for combined treatments, drugs were added together with 20% serum.

Antibodies, western blotting and immunofluorescence. Western blotting was carried out as in ref. 28. Immunofluorescence was as in ref. 29. Antibodies: anti-YAP/TAZ 1:200 for immunofluorescence (sc101199 detecting both YAP and TAZ in western blot), anti-YAP 1:100 for immunofluorescence (sc271134 detecting only YAP in western blot, used in Supplementary Fig. 5), anti-phosphoS127-YAP (CST 4911), anti-LATS1 (CST 3477), anti-LATS2 (Abnova ab70565), anti-GAPDH (Millipore mAb374), anti-vinculin[30,31] (VIN-11-5). Primary antibodies for immunofluorescence were incubated overnight in PBS with 0.1% Triton and 2% goat serum. Secondary antibodies were GAM Alexa488, GAM Alexa568 and GAR Alexa555 (Invitrogen). YOYO1, TOTO3 (Invitrogen) or Hoechst were used in combination with RNase to counterstain nuclei. Alexa 488-conjugated phalloidin (Invitrogen) was used 1:100 in 1% BSA to visualize F-actin microfilaments. Firm-setting anti-fade mounting medium was 10% Mowiol 4-88, 2.5% DABCO, 25% glycerol, 0.1 M Tris-HCl pH 8.5. Images were acquired with a Leica SP2 confocal microscope equipped with a CCD camera. Cells seeded on microposts were observed in 1× PBS with a Bio-Rad upright confocal microscope with water immersion long-range objectives. Pictures of cells seeded on small adhesive islands were rescaled to allow better visualization of immunostainings. For quantifications of YAP/TAZ subcellular localizations, YAP/TAZ immunofluorescence signal was scored as predominantly nuclear versus evenly distributed/predominantly cytoplasmic in 150–200 cells for each experimental condition.

Real-time PCR. Cultures were harvested in TRIzol (Invitrogen) for total RNA extraction, and contaminant DNA was removed by DNase I treatment. cDNA synthesis was carried out with dT-primed MuMLV Reverse Transcriptase (Invitrogen). Real-time qPCR analyses were carried out on triplicate samplings of retrotranscribed cDNAs with RG3000 Corbett Research thermal cycler and analysed with Rotor-Gene Analysis6.1 software. Expression levels are given relative to *GAPDH*. Sequences of primers are provided in Supplementary Table 3.

Biostatistical analysis. The statistical association between genes differentially expressed in mammary epithelial cells (MEC) cultivated on ECM of high/low stiffness (stiffness signature) and belonging to signal transduction pathways is assessed by an over-representation analysis approach using Fisher's exact test. Briefly, considering that there are S single-symbol-annotated genes on the stiffness signature, the over-representation of a pre-defined pathway signature is calculated as the hypergeometric probability of having α genes for a specific pathway in S, under the null hypothesis that they were picked out randomly from the N total genes of the microarray. Over-representation analysis has been conducted using one-sided Fisher's exact test (*phyper* function of R *stats* package; P-value $<$ 0.05) and considering 19,621 single-symbol-annotated genes on the HG-U133 Plus2.0 microarray. P-values have been adjusted for false discovery rate (*p.adjust* function of R *stats* package; FDR $<$ 5%).

The stiffness signature has been derived from Supplementary Table 1 of ref. 9. The complete signature contains 1,236 probe sets of the Affymetrix 430 2.0 mouse array accounting for 1,015 single-symbol-annotated. MOE430 Plus2.0 probe IDs have been converted to the correspondent HG-U133 Plus2.0 probe sets using the NetAffx orthologue annotation file derived from the NCBI HomologoGene database (MOE430A Orthologues/Homologues Release 30, http://www.affymetrix.com/). This conversion table allows mapping orthologous probe sets (that is, probe sets interrogating transcripts from orthologous genes) across two Affymetrix types of arrays. The 1,236 mouse probe sets of the stiffness signature were converted into 1,793 human probe sets corresponding to 807 single-symbol-annotated genes. Similarly, probe sets of all pathway signatures have been first converted into HG-U133 Plus2.0 probe sets, and then annotated as gene symbols using Bioconductor *hgu133plus2.db* package (release 2.3.5). Gene-sets of specific signalling pathways have been derived from: TGFβa[32]; TGFβb[33]; H-Ras and β-catenin[34]; ERBB2[35]; YAP[36–38]; YAP/TAZ[39]; WNT[40]; Notch and NICD[41]. The "YAP/TAZ signature" was published as supplemental table in ref. 39. The second "YAP signature" of Supplementary Fig. 2 is provided in Supplementary Table 1. See Supplementary Tables 4, 5 and 6 for the following signatures, that were derived from the microarrays published in: MAL/SRFa[42]; MAL/SRFb[43]; NF-κB[44]. Genes of WNT and β-Catenin pathway lists were not represented in the stiffness signature.

24. Tse, J. R. & Engler, A. J. Preparation of hydrogel substrates with tunable mechanical properties. *Curr. Protoc. Cell Biol.* **47,** 10.16.1–10.16.16 (2010).
25. du Roure, O. *et al.* Force mapping in epithelial cell migration. *Proc. Natl Acad. Sci. USA* **102,** 2390–2395 (2005).
26. Mahoney, W. M. Jr, Hong, J. H., Yaffe, M. B. & Farrance, I. K. The transcriptional co-activator TAZ interacts differentially with transcriptional enhancer factor-1 (TEF-1) family members. *Biochem. J.* **388,** 217–225 (2005).
27. Martello, G. *et al.* A MicroRNA targeting Dicer for metastasis control. *Cell* **141,** 1195–1207 (2010).
28. Dupont, S. *et al.* FAM/USP9x, a deubiquitinating enzyme essential for TGFβ signaling, controls Smad4 monoubiquitination. *Cell* **136,** 123–135 (2009).
29. Morsut, L. *et al.* Negative control of Smad activity by ectodermin/Tif1γ patterns the mammalian embryo. *Development* **137,** 2571–2578 (2010).
30. Galbraith, C. G., Yamada, K. M. & Sheetz, M. P. The relationship between force and focal complex development. *J. Cell Biol.* **159,** 695–705 (2002).
31. Giannone, G., Jiang, G., Sutton, D. H., Critchley, D. R. & Sheetz, M. P. Talin1 is critical for force-dependent reinforcement of initial integrin-cytoskeleton bonds but not tyrosine kinase activation. *J. Cell Biol.* **163,** 409–419 (2003).

32. Padua, D. *et al.* TGFβ primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133,** 66–77 (2008).
33. Adorno, M. *et al.* A mutant-p53/Smad complex opposes p63 to empower TGFβ-induced metastasis. *Cell* **137,** 87–98 (2009).
34. Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439,** 353–357 (2006).
35. Mackay, A. *et al.* cDNA microarray analysis of genes associated with *ERBB2* (HER2/*neu*) overexpression in human mammary luminal epithelial cells. *Oncogene* **22,** 2680–2688 (2003).
36. Zhao, B. *et al.* TEAD mediates YAP-dependent gene induction and growth control. *Genes Dev.* **22,** 1962–1971 (2008).
37. Dong, J. *et al.* Elucidation of a universal size-control mechanism in *Drosophila* and mammals. *Cell* **130,** 1120–1133 (2007).
38. Ota, M. & Sasaki, H. Mammalian Tead proteins regulate cell proliferation and contact inhibition as transcriptional mediators of Hippo signaling. *Development* **135,** 4059–4069 (2008).
39. Zhang, H. *et al.* TEAD transcription factors mediate the function of TAZ in cell growth and epithelial-mesenchymal transition. *J. Biol. Chem.* **284,** 13355–13362 (2009).
40. DiMeo, T. A. *et al.* A novel lung metastasis signature links Wnt signaling with cancer cell self-renewal and epithelial-mesenchymal transition in basal-like breast cancer. *Cancer Res.* **69,** 5364–5373 (2009).
41. Mazzone, M. *et al.* Dose-dependent induction of distinct phenotypic responses to Notch pathway activation in mammary epithelial cells. *Proc. Natl Acad. Sci. USA* **107,** 5012–5017 (2010).
42. Descot, A. *et al.* Negative regulation of the EGFR-MAPK cascade by actin-MAL-mediated Mig6/Errfi-1 induction. *Mol. Cell* **35,** 291–304 (2009).
43. Selvaraj, A. & Prywes, R. Expression profiling of serum inducible genes identifies a subset of SRF target genes that are MKL dependent. *BMC Mol. Biol.* **5,** 13 (2004).
44. Park, B. K. *et al.* NF-κB in breast cancer cells promotes osteolytic bone metastasis by inducing osteoclastogenesis via GM-CSF. *Nature Med.* **13,** 62–69 (2006).

doi:10.1038/nature10091

A current filamentation mechanism for breaking magnetic field lines during reconnection

H. Che[1], J. F. Drake[2] & M. Swisdak[3]

During magnetic reconnection, the field lines must break and reconnect to release the energy that drives solar and stellar flares[1,2] and other explosive events in space[3] and in the laboratory[4]. Exactly how this happens has been unclear, because dissipation is needed to break magnetic field lines and classical collisions are typically weak. Ion–electron drag arising from turbulence[5], dubbed 'anomalous resistivity', and thermal momentum transport[6] are two mechanisms that have been widely invoked. Measurements of enhanced turbulence near reconnection sites in space[7,8] and in the laboratory[9,10] support the anomalous resistivity idea but there has been no demonstration from measurements that this turbulence produces the necessary enhanced drag[11]. Here we report computer simulations that show that neither of the two previously favoured mechanisms controls how magnetic field lines reconnect in the plasmas of greatest interest, those in which the magnetic field dominates the energy budget. Rather, we find that when the current layers that form during magnetic reconnection become too intense, they disintegrate and spread into a complex web of filaments that causes the rate of reconnection to increase abruptly. This filamentary web can be explored in the laboratory or in space with satellites that can measure the resulting electromagnetic turbulence.

Particle-in-cell simulations reveal that the rate of reconnection in a three-dimensional system jumps sharply above the rate measured in two dimensions (Fig. 1a). The jump is a consequence of the development of turbulence in three dimensions. To understand why this happens, it is necessary to explore how magnetic fields break and reconnect to form the topological x-line that characterizes reconnection (Fig. 2a). Magnetic reconnection produces large electric fields that are parallel to the local magnetic field in the vicinity of the x-line (Fig. 2b). In the absence of a dissipative process, such a parallel electric field would produce an infinite current and reconnection would cease. The mechanisms that limit the electron response to parallel electric fields therefore break field lines. These mechanisms can be understood from the electron momentum equation in the direction perpendicular to the plane of reconnection (the z direction)[12]:

$$\frac{\partial p_{ez}}{\partial t} + \nabla\cdot\mathbf{p}_e v_{ez} = -en_eE_z + \frac{1}{c}(\mathbf{j}_e \times \mathbf{B})_z - \nabla\cdot\mathbf{P}_{ez} \qquad (1)$$

where $\mathbf{p}_e$ is the momentum density, $\mathbf{j}_e$ is the electron current density and $\mathbf{P}_{ez} = P_e\cdot\hat{z}$ with P_e the pressure tensor. In a three-dimensional system where the intense current layers produced during reconnection (Fig. 2c) drive short-scale turbulence with wavevectors along z, the impact of this turbulence can be quantified by averaging equation (1) over z to obtain a generalized Ohm's law:

$$<E_z> = -\frac{1}{c}(<\mathbf{v}_{e\perp}> \times <\mathbf{B}_\perp>)_z - \frac{\nabla_\perp\cdot<\mathbf{P}_{ez}>}{<n_e>e}$$
$$-\frac{m_e}{e}\left(\frac{\partial<v_{ez}>}{\partial t} + <\mathbf{v}_{e\perp}>\cdot\nabla_\perp\cdot<v_{ez}>\right) + D_{ez} + \nabla_\perp\cdot\mathbf{T}_{ez}$$

where $<f>$ denotes an average of f over z with $\delta f = f - <f>$, and the subscript $\perp$ denotes the x–y plane.

$$D_{ez} = \frac{-1}{<n_e>}<\delta n_e \delta E_z>$$

and

$$\mathbf{T}_{ez} = -\frac{1}{<n_e>e}<\delta\mathbf{p}_{e\perp}\left(\delta v_{ez} - \frac{e}{m_ec}\delta A_z\right)>$$

are the turbulent drag and momentum transport, respectively, and δA_z is the vector potential for $\delta\mathbf{B}_\perp$. This representation for $\mathbf{T}_{ez}$ is valid only in the strong guide field limit (see the Supplementary Information). In a two-dimensional system D_{ez} and $\mathbf{T}_{ez}$ are both zero and $\mathbf{v}_{e\perp}$ is typically zero at the x-line. In steady state the inertia term is also zero, so only the thermal momentum transport arising from the xz and yz components of $\mathbf{P}_{ez}$ can balance the reconnection electric field. This was confirmed in earlier two-dimensional simulations when the ambient guide magnetic field B_z is small or absent[6]. However, in systems in which the magnetic energy exceeds the local electron plasma pressure so that $\beta_e = 8\pi n_eT_e/B^2$ is small, such as in solar and stellar coronae and some regions of the Earth's magnetosphere, we show that thermal momentum transport is weak. Therefore, in such systems other mechanisms for balancing the reconnection electric field need to be explored.

In a three-dimensional system the strong currents driven by the reconnection electric field around the x-line could potentially drive turbulence, scattering electrons to produce an electron–ion drag or anomalous resistivity to balance the reconnection electric field[5,12–14]. This turbulence is typically weak or absent in two dimensions because the growth rates of instabilities peak when wavevectors have components along the direction of the current, which is strongest along z. The drag D_{ez} in equation (2) measures the effect of this turbulence but neither simulations nor observations have established the viability of the mechanism. The narrow current layers that result from reconnection can also break up through self-driven turbulence. The resulting turbulent transport of electron momentum away from the x-line as described by $\mathbf{T}_{ez}$ could balance the reconnection electric field and act to break field lines[15,16]. We explore this idea below.

At larger spatial scales ambient turbulence, which is common in astrophysical plasmas, might enable fast reconnection by breaking up the large-scale current layers that normally characterize magnetohydrodynamic reconnection[17,18]. In such models, however, the question of what provides the dissipation required to break field lines is not addressed.

The development of turbulence during simulations of three-dimensional reconnection and its consequences are presented in Figs 1–4. A comparison of the time-development of the rate of reconnection between two and three dimensions reveals that for a system that initially has low β_e the rate of reconnection rises over time, with the two simulations matching until the rate in the three-dimensional case abruptly increases (Fig. 1a). Coincident with the rate increase in three dimensions is an increase in the turbulent-driven drag and momentum transport (Fig. 1b). The reconnection electric field E_z drives a strong current around the x-line that strengthens and narrows over time (Fig. 1a). At late time the current layer in the three-dimensional case

[1]Center For Integrated Plasma Studies, Department of Physics, University of Colorado, Boulder, Colorado 80309, USA. [2]Department of Physics and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA. [3]Department of Physics, University of Maryland, College Park, Maryland 20742, USA.

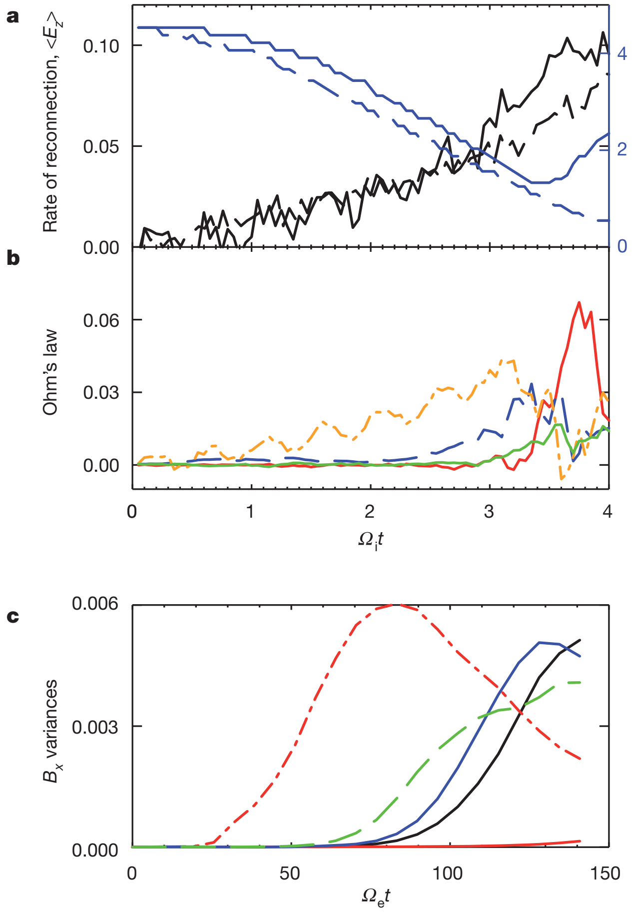


Figure 1 | The time evolution of reconnection and the development of turbulence. Particle-in-cell simulations using the p3d code[27] are performed in doubly periodic two-dimensional and three-dimensional geometries starting with two force-free current sheets. The reconnection magnetic field is $B_x/B_0 = \tanh[(y - L_y/4)/w_0] - \tanh[(y - 3L_y/4)/w_0] - 1$, where the current layer width $w_0 = 0.5d_i$ and the box size is $L_x \times L_y \times L_z = 4d_i \times 2d_i \times 8d_i$. The electron and ion temperatures, $T_e/m_i c_A{}^2 = T_i/m_i c_A{}^2 = 0.04$, and density n_0 are initially uniform. The initial out-of-plane 'guide' field B_z/B_0 is 5.0 outside the current layers and increases within the current layers so that the total magnetic field B is a constant. The cyclotron time is $\Omega_i^{-1} = m_i c/eB_0$, the Alfvén speed is $c_A = B_0/(4\pi m_i n_0)^{1/2}$ and the ion inertial length is $d_i = c_A/\Omega_i$. The electron mass is $0.01m_i$ and the velocity of light $c = 20c_A$. The three-dimensional spatial grid consists of $512 \times 256 \times 1{,}024$ cells with 20 particles per cell, whereas in two dimensions it consists of $2{,}048 \times 1{,}024$ with 100 particles per cell. Reconnection is initiated with a large-scale magnetic perturbation. In **a** the rates of reconnection $<E_z>$ (normalized to $E_0 = c_A B_0/c$) and the half-widths of the z-averaged electron current layer at the x-line (see Fig. 2a) in three dimensions (solid lines) and two dimensions (dashed lines) are shown. In **b** the dominant components of Ohm's law (normalized to E_0), the electron inertia $-(m_e/e)\partial<v_{ez}>/\partial t$ (orange line), the drag D_z (blue line), the y-directed turbulent momentum transport $\partial T_{yz}/\partial y$ (red line) and the thermal momentum transport $-\nabla_\perp \cdot <\mathbf{P}_{ez}>/<n_e>e$ (green line) are shown. In three-dimensional simulations with four times the initial electron and ion temperatures (not shown), no turbulence develops and thermal momentum transport balances $<E_z>$. Finally, in **c** the magnetic field variances $<\delta B_x>^2/B_0{}^2$ are plotted versus time from a series of three-dimensional simulations of narrow current layers (no reconnection) with various guide fields B_z/B_0 (5.0 in black, 2.5 in green, 1.0 in blue and 0.5 in red) and stationary ions are shown. The simulation domains are $L_x \times L_y \times L_z = 10d_e \times 10d_e \times 80d_e$ and all but the dot-dashed red line have current layer widths of d_e. All but the guide field of 0.5 break up, but reducing the layer width in this case from $1.0d_e$ to $0.5d_e$ (dot-dashed red line) again leads to break-up. Current layers with Harris equilibria also break up (see the Supplementary Information).

abruptly broadens as the nearly laminar current layer undergoes a transition to a filamented current layer (Fig. 3a and b). The sharp jump in the strength of the magnetic field perturbations (Fig. 3c and d) and

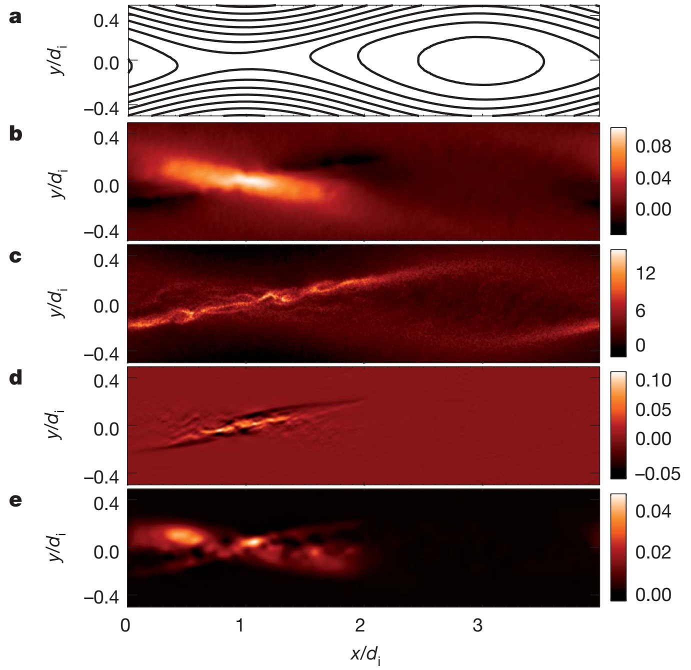


Figure 2 | The geometry of magnetic reconnection at late time. Shown is the x–y plane at $t = 3.75\ \Omega_{ci}^{-1}$ from the three-dimensional simulation of Fig. 1. In **a** the reconnected magnetic field lines (averaged over z) are shown. In **b** the field line motion towards and away from the x-line induces an electric field that produces the parallel (to **B**) electric field $<\mathbf{E}_\parallel>$ that drives the intense electron current J_{ez} around the x-line in **c**. The irregular structure of the current layer indicates that it is turbulent (see also Fig. 3). The filamentation of the current layer transports electron momentum p_{ez} away from the centre of the current layer. The rate of momentum transport $\nabla \cdot \mathbf{T}_{ez}$ is shown in **d**. The positive peak in $\nabla \cdot \mathbf{T}_{ez}$ with negative values upstream demonstrates that momentum is being transported away from the x-line. The turbulence also produces a net electron–ion drag D_z, shown in **e**. We note that the spatial structure of the drag is very different from the turbulent momentum transport. It does not overlap with the region of high current density and therefore cannot be represented as an anomalous resistivity. The colour scales measure the variables normalized to E_0 (**b**, **d** and **e**) and $n_0 e c_A$ (**c**).

the wavelength of the electric field perturbations (Fig. 3e and f) indicate that this filamentation instability is distinct from instabilities explored earlier in observations[7,8] and modelling[19–23].

To explore the role of this turbulence in reconnection and determine whether it facilitates the breaking of field lines, we evaluate Ohm's law in equation (2) around the x-line (Fig. 4). In two dimensions the reconnection electric field is balanced by the electron inertia even at late time. In three dimensions the mechanisms controlling how the magnetic field lines break evolve over time: early in time it is the electron inertia, at intermediate times the drag reaches parity with the inertia, and at late time the turbulent momentum transport dominates. The broadening of the current layer (Fig. 1a) causes the electron streaming velocity, the inertia term in Ohm's law and the turbulent drag to decrease sharply at the x-line (Fig. 1b). By the end of the simulation the momentum transport has also decreased because of the broadening of the current layer. In larger simulation domains, where more magnetic flux is available to reconnect, the process of current layer thinning and broadening goes through several cycles before coming to an approximate balance (see the movie in the Supplementary Information).

What drives the filamentation of the current shown in Fig. 3? The time-dependence of the transverse electric fields reveals that it is a right-hand circularly polarized electromagnetic wave and hence is part of the whistler/electron-cyclotron branch. The spatial correlation between the current density and transport in Fig. 2 suggests that it is driven by the current density gradient and not by the relative streaming velocity of electrons and ions. Three-dimensional particle-in-cell

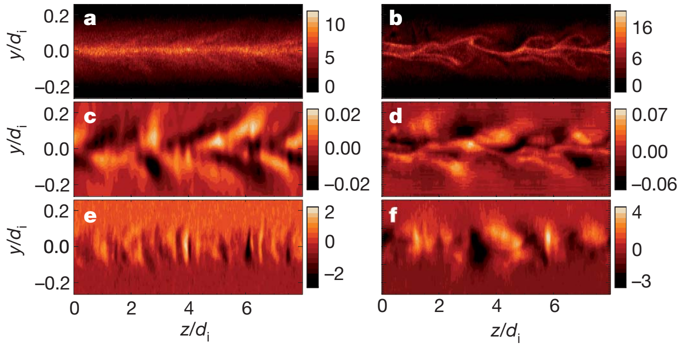


Figure 3 | The filamentary structure of the electron current layer. In the y–z plane in a cut through the x-line at $t=3.0\Omega_{ci}^{-1}$ (**a**, **c** and **e**) and $t=3.75\Omega_{ci}^{-1}$ (**b**, **d** and **f**) are the electron current j_{ez}/n_0ec_A (**a** and **b**), the magnetic field perturbations $\delta B_x/B_0$ (**c** and **d**) and the electric field perturbations $\delta E_x/E_0$ (**e** and **f**). The filamentation instability begins sharply at $t=3.5\Omega_{ci}^{-1}$. This instability did not appear in earlier simulations[13,14,28] because the computational domains along the z axis were too short to capture its long wavelength parallel to $\boldsymbol{B}$ (in ref. 28 the computational domain along the z axis calculated from the upstream density was only $2.86d_i$). The measured phase speed of the filaments is close to the electron drift speed, suggesting that the coupling to the ions is weak. The colour scales measure the normalized variables.

simulations of narrow current layers reveal that the filamentation instability is insensitive to the ion mass and remains robust even for stationary ions (see Supplementary Information). A simple fluid description of electromagnetic waves in this regime is given by the electron magnetohydrodynamic equations[24] and the possibility that gradients in the electron current could drive instability has been previously discussed[16,25]. Linearization of the electron magnetohydrodynamic

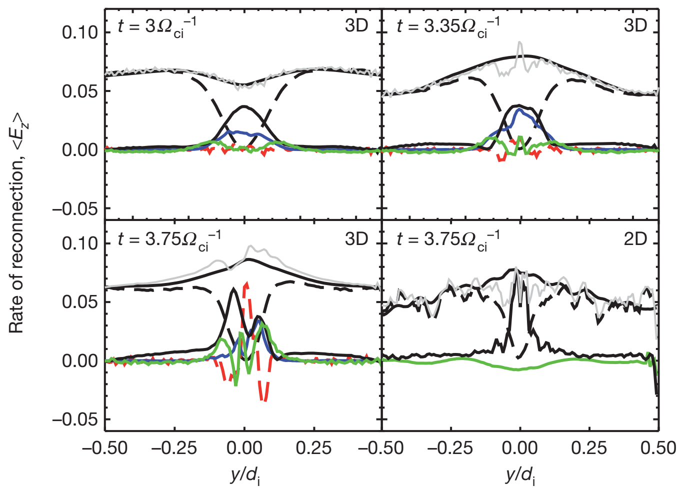


Figure 4 | Breaking magnetic field lines: the dominant components of Ohm's law. At $t=3.0\Omega_{\rm ci}^{-1}$, $t=3.35\Omega_{\rm ci}^{-1}$ and $t=3.75\Omega_{\rm ci}^{-1}$ from the three-dimensional (3D) simulation and $t=3.75\,\Omega_{\rm ci}^{-1}$ from the two-dimensional (2D) simulation, the dominant contributions to Ohm's law are plotted in cuts along the inflow direction y through the x-line. The lines shown are $\langle E_z\rangle$ (upper solid black), $-\langle v_{ey}\rangle\langle B_x\rangle/c$ (dashed black), electron inertia (lower solid black), the thermal momentum transport $-(1/\langle n_e\rangle e)\partial\langle P_{eyz}\rangle/\partial y$ (solid green), the drag D_z (blue) and turbulent transport $\partial T_{yz}/\partial y$ (red). The solid grey line shows the sum of all of the contributions to Ohm's law, which should lie on top of $\langle E_z\rangle$. In two dimensions the electron inertia continues to balance the reconnection electric field at late time, a solution that is not consistent with steady reconnection. In three dimensions the turbulent drag and inertia dominate at $t=3.0\Omega_{\rm ci}^{-1}$. By $t=3.35\Omega_{\rm ci}^{-1}$ the turbulent drag and the rate of reconnection both sharply increase, suggesting a causal relation. By $t=3.75\Omega_{\rm ci}^{-1}$, the current layer is becoming filamentary and turbulent momentum transport completely dominates the force balance at the x-line. Momentum is transported upstream away from the x-line, producing a positive $\partial T_{yz}/\partial y$ at the x-line and negative values upstream corresponding to a momentum transfer, not a momentum sink. At the x-line the drag drops sharply as the dynamics of the filaments dominates.

equations in the presence of a local current density gradient $J_{ez}'=\mathrm{d}J_{ez}/\mathrm{d}y$ yields an instability with a peak growth rate of $\gamma\approx J_{ez}'/(2ne)$ (see Supplementary Information). Of course, a kinetic treatment will be required to understand this instability fully.

What is the range of parameters (for example, the guide field) over which the instability is important? A series of three-dimensional particle-in-cell simulations of narrow current layers reveal a surprising result: the guide field has a destabilizing rather than a stabilizing influence (Fig. 1c). The instability requires a finite guide field to develop, is strongest for a guide field that is around twice the reconnecting magnetic field, but remains robust down to a guide field of 0.5 of the reconnecting field for current layer widths of $0.5d_e$, where d_e is the electron skin depth (see also the Supplementary Information). During reconnection in a two-dimensional system with a guide field, the width of the current layer decreases to the electron Larmor radius $\rho_e=\beta_e^{0.5}d_e$ (refs 6 and 26). Thus, in the Earth's magnetosphere, reconnection-driven current layers should filament for values of β_e below 0.25 and guide fields around 0.5.

Satellites such as Cluster and THEMIS should be able to measure the predicted asymmetry in the distribution of turbulence between the two separatrices (Fig. 2c, d). Detailed measurements of d_e-scale current layers must await the results of the Magnetospheric Multiscale Mission. Filamentation of current layers in the Versatile Toroidal Facility reconnection experiment at the Massachusetts Institute of Technology has already been observed (J. Egedal, personal communication).

Received 22 June 2010; accepted 1 April 2011.
Published online 1 June 2011.

1. Tsuneta, S. Heating and acceleration processes in hot thermal and impulsive solar flares. *Astrophys. J.* **290,** 353–358 (1985).
2. Priest, E. R. & Forbes, T. G. *Magnetic Reconnection: MHD Theory and Applications* (Cambridge University Press, 2000).
3. Baker, D. N., Pulkkinen, T. I., Angelopoulos, V., Baumjohann, W. & McPherron, R. L. Neutral line model of substorms: past results and present view. *J. Geophys. Res.* **101,** 12975–13010 (1996).
4. Yamada, M. *et al.* Investigation of magnetic reconnection during a sawtooth crash in a high-temperature tokamak plasma. *Phys. Plasmas* **1,** 3269–3276 (1994).
5. Galeev, A. A. & Sagdeev, R. Z. in *Basic Plasma Physics* (eds Galeev, A. A. & Sudan, R. N.) Vol. 1, 677–731 (North Holland Publishing Company, 1983).
6. Hesse, M., Kuznetsova, M. & Hoshino, M. The structure of the dissipation region for component reconnection: particle simulations. *Geophys. Res. Lett.* **29,** 1563, doi:10.1029/2001GL014714 (2002).
7. Matsumoto, H., Deng, X. H., Kojima, H. & Anderson, R. R. Observation of electrostatic solitary waves associated with reconnection on the dayside magnetopause boundary. *Geophys. Res. Lett.* **30,** 1326, doi:10.1029/2002GL016319 (2003).
8. Cattell, C. *et al.* Cluster observations of electron holes in association with magnetotail reconnection and comparison to simulations. *J. Geophys. Res.* **110,** A01211 (2005).
9. Ji, H. *et al.* Electromagnetic fluctuations during fast reconnection in a laboratory plasma. *Phys. Rev. Lett.* **92,** 115001 (2004).
10. Fox, W. *et al.* Laboratory observation of electron phase-space holes during magnetic reconnection. *Phys. Rev. Lett.* **101,** 255003 (2008).
11. Eastwood, J., Phan, T. D., Bale, S. D. & Tjulin, A. Observations of turbulence generated by magnetic reconnection. *Phys. Rev. Lett.* **102,** 035001 (2009).
12. Vasyliunas, V. M. Theoretical models of magnetic field line merging. 1. *Rev. Geophys. Space Phys.* **13,** 303–336 (1975).
13. Drake, J. F. *et al.* Formation of electron holes and particle energization during magnetic reconnection. *Science* **299,** 873–877 (2003).
14. Che, H., Drake, J. F., Swisdak, M. & Yoon, P. H. Electron holes and heating in the reconnection dissipation region. *Geophys. Res. Lett.* **37,** L11105, doi:10.1029/2010GL043608 (2010).
15. Kaw, P. K., Valeo, E. J. & Rutherford, P. H. Tearing modes in a plasma with magnetic braiding. *Phys. Rev. Lett.* **43,** 1398–1401 (1979).
16. Drake, J. F., Kleva, R. G. & Mandt, M. E. Structure of thin current layers: implications for magnetic reconnection. *Phys. Rev. Lett.* **73,** 1251–1254 (1994).
17. Lazarian, A. & Vishniac, E. Reconnection in a weakly stochastic field. *Astrophys. J.* **517,** 700, doi:10.1086/307233 (1999).
18. Kowal, G., Lazarian, A., Vishniac, E. & Otmianowska-Mazur, K. Reconnection in a weakly stochastic magnetic field. *Astrophys. J.* **700,** 63–85 (2009).
19. Openheim, M. M., Vetoulis, G., Newman, D. L. & Goldman, M. V. Evolution of electron phase-space holes in 3-D. *Geophys. Res. Lett.* **28,** 1891–1894 (2001).
20. Omura, Y., Matsumoto, H., Miyake, T. & Kojima, H. Electron beam instabilities as a generation mechanism of electrostatic solitary waves in the magnetotail. *J. Geophys. Res.* **101,** 2685–2697 (1996).
21. McMillan, B. F. & Cairns, I. H. Lower hybrid turbulence driven by parallel currents and associated electron energization. *Phys. Plasmas* **13,** 052104, doi:10.1063/1.2198212 (2006).

22. Goldman, M. V., Newman, D. L. & Pritchett, P. Vlasov simulations of electron holes driven by particle distribution from PIC reconnection simulations with a guide field. *Geophys. Res. Lett.* **35,** L22109, doi:10.1029/2008GL035608 (2008).
23. Che, H., Drake, J. F., Swisdak, M. & Yoon, P. H. Nonlinear development of streaming instabilities in strongly magnetized plasma. *Phys. Rev. Lett.* **102,** 145004 (2009).
24. Kingsep, A. S. Chukbar, K. V. & Yan'kov, Y. Y. in *Reviews of Plasma Physics* (ed. Kadomtsev, B. B.) Vol. 16, 243–288 (Consultants Bureau, 1990).
25. Ferraro, N. M. & Rogers, B. N. Turbulence in low-β reconnection. *Phys. Plasmas* **11,** 4382–4389 (2004).
26. Ricci, P. Brackbill, J. U., Daughton, W. & Lapenta, G. Collisionless magnetic reconnection in the prescence of a guide field. *Phys. Plasmas* **11,** 4102–4114 (2004).
27. Zeiler, A. *et al.* Three-dimensional particle simulations of collisionless magnetic reconnection. *J. Geophys. Res.* **107,** 1230, doi:10.1029/2001JA000287 (2002).
28. Pritchett, P. & Coroniti, F. V. Three-dimensional collisionless magnetic reconnection in the presence of a guide field. *J. Geophys. Res.* **109,** A01220, doi:10.1029/2003JA009999 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work has been supported by the NSF/DOE programme in plasma science and by NASA through the Supporting Research and Technology Program and the Magnetospheric Multiscale Mission Science Team. Computations were carried out in part at the National Energy Research Scientific Computing Center.

Author Contributions All of the authors made significant contributions to this work. H.C. carried out the particle simulations of reconnection. M.S. carried out simulations of isolated electron current layers. H.C., J.F.D. and M.S. analysed the data from the simulations. All authors discussed the results and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.C. (haihong.che@colorado.edu).

doi:10.1038/nature10120

Direct measurement of the quantum wavefunction

Jeff S. Lundeen[1], Brandon Sutherland[1], Aabid Patel[1], Corey Stewart[1] & Charles Bamber[1]

The wavefunction is the complex distribution used to completely describe a quantum system, and is central to quantum theory. But despite its fundamental role, it is typically introduced as an abstract element of the theory with no explicit definition[1,2]. Rather, physicists come to a working understanding of the wavefunction through its use to calculate measurement outcome probabilities by way of the Born rule[3]. At present, the wavefunction is determined through tomographic methods[4–8], which estimate the wavefunction most consistent with a diverse collection of measurements. The indirectness of these methods compounds the problem of defining the wavefunction. Here we show that the wavefunction can be measured directly by the sequential measurement of two complementary variables of the system. The crux of our method is that the first measurement is performed in a gentle way through weak measurement[9–18], so as not to invalidate the second. The result is that the real and imaginary components of the wavefunction appear directly on our measurement apparatus. We give an experimental example by directly measuring the transverse spatial wavefunction of a single photon, a task not previously realized by any method. We show that the concept is universal, being applicable to other degrees of freedom of the photon, such as polarization or frequency, and to other quantum systems—for example, electron spins, SQUIDs (superconducting quantum interference devices) and trapped ions. Consequently, this method gives the wavefunction a straightforward and general definition in terms of a specific set of experimental operations[19]. We expect it to expand the range of quantum systems that can be characterized and to initiate new avenues in fundamental quantum theory.

The wavefunction Ψ, also known as the 'quantum state', is considerably more difficult to measure than the state of a classical particle, which is determined simply by measuring its position X and momentum P. According to the Heisenberg uncertainty principle, in quantum theory a precise measurement of X disturbs the particle's wavefunction and forces a subsequent measurement of P to become random. Thus we learn nothing of the particle's momentum. Indeed, it is impossible to determine a completely unknown wavefunction of single system[20].

Consider instead performing a measurement of X on an ensemble of particles, all with the same Ψ. The probability of getting result $X=x$ is $|\Psi(x)|^2$. Similarly, the probability of $P=p$ would be $|\Phi(p)|^2$, where $\Phi(p)$ is the Fourier transform of $\Psi(x)$. Even these two probability distributions are not enough to determine $\Psi(x)$ unambiguously (see the one-dimensional phase retrieval problem[21]). Instead, Ψ must be reconstructed by performing a large set of distinct measurements (for example, of quadratures $Q(\theta)=X\cos(\theta)+P\sin(\theta)$, for mixing angles, θ, ranging from 0 to 2π), and then estimating a Ψ that is most compatible with the measurement results. This method is known as quantum state tomography[4–8]. In contrast, we introduce a method to measure Ψ of an ensemble directly. By 'direct' we mean that the method is free from complicated sets of measurements and computations; the average raw signal originating from where the wavefunction is being probed is simply proportional to its real and imaginary components at that point. The method rests on the sequential measurement of two complementary variables of the system.

At the centre of the direct measurement method is a reduction of the disturbance induced by the first measurement. Consider the measurement of an arbitrary variable A. In general, measurement can be seen as the coupling between an apparatus and a physical system that results in the translation of a pointer. The pointer position indicates the result of a measurement. In a technique known as 'weak measurement', the coupling strength is reduced and this correspondingly reduces the disturbance created by the measurement[9–18]. This strategy also compromises measurement precision, but this can be regained by averaging. The average of the weak measurement is simply the expectation value $\langle\Psi|A|\Psi\rangle$, indicated by an average position shift of the pointer proportional to this amount.

A distinguishing feature of weak measurement is that it does not disturb a subsequent normal (or 'strong') measurement of another observable C in the limit where the coupling vanishes. For the particular ensemble subset that gave outcome $C=c$, one can derive the average of the weak measurement of A. In the limit of zero interaction strength, this is called the 'weak value' and is given[9] by:

$$\langle A\rangle_{\mathrm{W}}=\frac{\langle c|A|\Psi\rangle}{\langle c|\Psi\rangle} \tag{1}$$

Selecting a particular subset of an ensemble based on a subsequent measurement outcome is known as 'post-selection', and is a common tool in quantum information processing[22,23].

Unlike the standard expectation value $\langle A\rangle$, the weak value $\langle A\rangle_{\mathrm{W}}$ can be a complex number. This seemingly strange result can be shown to have a simple physical manifestation: the pointer's position is shifted by $\mathrm{Re}\langle A\rangle_{\mathrm{W}}$ and receives a momentum kick of $\mathrm{Im}\langle A\rangle_{\mathrm{W}}$ (refs 24–26). The complex nature of the weak value suggests that it could be used to indicate both the real and the imaginary parts of the wavefunction.

Returning to our example of a single particle, consider the weak measurement of position ($A=\pi_x\equiv|x\rangle\langle x|$) followed by a strong measurement of momentum giving $P=p$. In this case, the weak value is:

$$\langle\pi_x\rangle_{\mathrm{W}}=\frac{\langle p|x\rangle\ \langle x|\Psi\rangle}{\langle p|\Psi\rangle} \tag{2}$$

$$=\frac{e^{ipx/\hbar}\Psi(x)}{\Phi(p)} \tag{3}$$

In the case $p=0$, this simplifies to

$$\langle\pi_x\rangle_{\mathrm{W}}=k\Psi(x) \tag{4}$$

where $k=1/\Phi(0)$ is a constant (which can be eliminated later by normalizing the wavefunction). The average result of the weak measurement of π_x is proportional to the wavefunction of the particle at x. Scanning the weak measurement through x gives the complete wavefunction. At each x, the observed position and momentum shifts of the measurement pointer are proportional to $\mathrm{Re}\Psi(x)$ and $\mathrm{Im}\Psi(x)$, respectively. In short, by reducing the disturbance induced by measuring X and then measuring P normally, we measure the wavefunction of the single particle.

As an experimental example, we performed a direct measurement of the transverse spatial wavefunction of a photon. Considering a photon travelling along the z direction, we directly measure the x wavefunction

[1]Institute for National Measurement Standards, National Research Council, 1200 Montreal Road, Ottawa, Canada, K1A 0R6.

of the photon, sometimes called the 'spatial mode' (see Supplementary Discussion). The Wigner function of the spatial mode of a classical beam has been measured directly but not for a single photon state[27,28].

We produce a stream of photons in one of two ways, either by attenuating a laser beam or by generating single photons through spontaneous parametric down-conversion (SPDC; see Supplementary Methods for details). The photons have a centre wavelength of $\lambda = 783$ nm or 800 nm, respectively. The experiment (details and diagram in Fig. 1) can be divided into four sequential steps: preparation of the transverse wavefunction, weak measurement of the transverse position of the photon, post-selection of those photons with zero transverse momenta, and readout of the weak measurement.

An ensemble of photons with wavefunction $\Psi(x)$ is emitted from a single mode fibre and collimated. We begin by directly measuring this wavefunction (described in detail in Fig. 1). We then further test our method by inducing known magnitude and phase changes to the photons here to prepare a series of modified wavefunctions.

We weakly measure the transverse position of the photon by coupling it to an internal degree of freedom of the photon, its polarization. This allows us to use the linear polarization angle of the photon as the pointer. At a position x where we wish to measure $\pi_x = |x\rangle\langle x|$, we rotate the linear polarization of the light by α. Consider if α is set to 90°. In this case, one can perfectly discriminate whether a photon had position x because it is possible to perfectly discriminate between orthogonal polarizations, 0° and 90°. This is a strong measurement. Reducing the strength of the measurement corresponds to reducing α, which makes it impossible to discriminate with certainty whether any particular photon had $X = x$. The benefit of this reduction in precision is a commensurate reduction in the disturbance to the wavefunction of the single photon.

We then use a Fourier transform lens and a slit to post-select only those photons with $p = 0$. This constitutes the strong measurement of P.

In this subset of the photon ensemble, we find the average value of our weak measurement of π_x. The average rotation of the pointer, the linear polarization, is proportional to the real part of the weak value. Its complementary pointer variable, the rotation in the circular polarization basis, is proportional to the imaginary part of the weak value[25]. Formally, if we treat the initial polarization as a spin-1/2 spin-down vector, then the weak value is given by

$$\langle\pi_x\rangle_{\mathrm{W}} = \frac{1}{\sin\alpha}\left(\langle s|\sigma_x|\rangle s - i\langle s|\sigma_y|s\rangle\right) \tag{5}$$

where σ_x and σ_y are the Pauli x and y matrices, respectively, and $|s\rangle$ is the final polarization state of the pointer[25]. We measure the σ_x and σ_y expectation values by sending the photons through a half-wave plate or a quarter-wave plate, respectively, and then through a polarizing beamsplitter (PBS). Thus, we read out $\mathrm{Re}\Psi(x)$ (half-wave plate) and $\mathrm{Im}\Psi(x)$ (quarter-wave plate) from the signal imbalance between detectors 1 and 2 at the outputs of the PBS (Fig. 1).

With $\alpha = 20°$, we scan our measurement of π_x in 1-mm steps and find the weak value $\langle\pi_x\rangle_{\mathrm{W}}$ at each step. In this way, we directly measure the photon transverse wavefunction, $\Psi(x) = |\Psi(x)|\exp(i\phi(x))$. We normalize the σ_x and σ_y measurements by the same factor, so that $\int|\Psi(x)|^2\mathrm{d}x = 1$, which eliminates the proportionality constant, $\sin\alpha/\Phi(0)$.

To confirm our direct measurement method, we test it on a series of different wavefunctions. Using our SPDC single photon source, we start by measuring the initial truncated Gaussian wavefunction (Fig. 2) described in Fig. 1. Switching to the laser source of photons, we then modify the magnitude, and then the phase, of the initial wavefunction with an apodized filter and glass plate, respectively, to create two new test Ψ (Fig. 3). We conduct more quantitative modification of the wavefunction phase by introducing a series of phase gradients and then phase curvatures (Fig. 4). For all the test wavefunctions, we have found

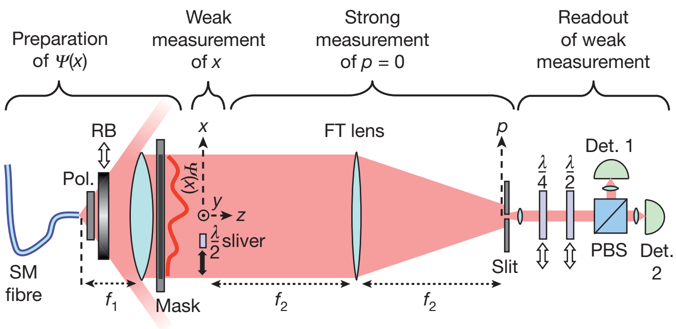


Figure 1 | Direct measurement of the photon transverse wavefunction. To begin with photons having identical wavefunctions, we transmit them through an optical fibre (Nufern PM780-HP) that allows only a single mode (SM) to pass. This mode is approximately Gaussian, with a nominal $1/e^2$ diameter of 5.3 ± 1.0 µm. The photons emerge from the fibre and pass through a micro-wire polarizer (Pol.; Edmund Optic NT47-602) to be collimated by an achromatic lens ($f_1 = 30$ cm, diameter 5 cm, Thorlabs AC508-300-B), one focal length (f_1) away from the fibre. The lens was masked off with a rectangular aperture of dimension $x \times y = 43$ mm $\times$ 11 mm. Thus our nominal initial wavefunction was a truncated Gaussian with a $1/e^2$ diameter of 56.4 mm and a flat phase profile. We modify the magnitude and phase of the nominal $\Psi(x)$ to create a series of test wavefunctions (see Figs 3, 4). At 45 mm past the lens, a rectangular sliver of a half-wave plate ($\lambda/2$ sliver) ($x \times y \times z$ dimensions of 1 mm $\times$ 25 mm $\times$ 1 mm) at position x is used to weakly measure $\pi_x = |x\rangle\langle x|$ (see Supplementary Methods for details). The photons then undergo an optical Fourier transform (FT) induced by an achromatic lens ($f_2 = 1$ m, diameter 5 cm, Thorlabs AC508-1000-B), placed one focal length (f_2) from the waveplate sliver. In the Fourier transform plane, one focal length f_2 past the lens, we post-select those photons with $p = 0$ by accepting only those that pass through a 15-µm-wide slit on axis. We collimate the photons emerging from the slit with an $f_3 = 3$ cm focal length lens. The photons pass through either a half-wave plate ($\lambda/2$) or quarter-wave plate ($\lambda/4$) and then through a polarizing beamsplitter (PBS). At each output port, the photons are focused onto a detector (Det. 1 and Det. 2): for the single photons, a photon counter (silicon avalanche photodiodes, PerkinElmer SPCM-AQHR-14); and for the laser, a silicon photodiode (Thorlabs, DET10A). The imbalance in counts or signal between the two detectors is proportional to the real ($\lambda/2$) or imaginary ($\lambda/4$) part of the wavefunction.

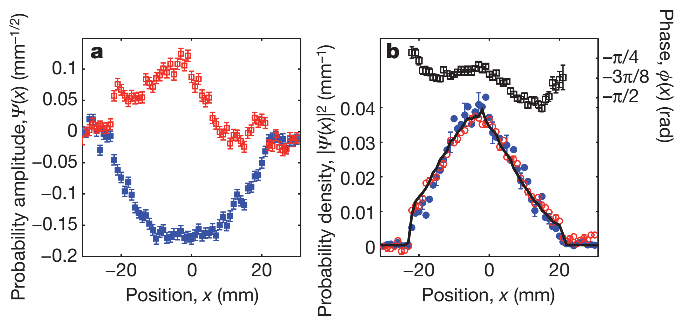


Figure 2 | The measured single-photon wavefunction, $\Psi(x)$, and its modulus squared and phase. **a**, $\mathrm{Re}\Psi(x)$ (solid blue squares) and $\mathrm{Im}\Psi(x)$ (open red squares) measured for the truncated Gaussian wavefunction. **b**, Using the data in **a** we plot the phase $\phi(x) = \arctan(\mathrm{Re}\Psi(x)/\mathrm{Im}\Psi(x))$ (open squares; right axis) and the modulus squared $|\Psi(x)|^2$ (solid blue circles; left axis). There is good agreement between the latter and a strong measurement of the x probability distribution Prob(x) (solid line; left axis) conducted by scanning a detector along x in the plane of the sliver. The phase is relatively flat, as expected from the fibre mode. The slight variation is consistent with the manufacturer specification of the first lens and the phase curvature measured with a shear plate. We also removed the slit completely. In this case, there is no post-selection and the weak value $\langle\pi_x\rangle$ becomes equal to the standard expectation value $\langle\Psi|\pi_x|\Psi\rangle = |\Psi(x)|^2$. We plot the measured $\mathrm{Re}\langle\pi_x\rangle$ (open red circles; left axis) after it is normalized so that $\int\mathrm{Re}\,\Psi(x)\mathrm{d}x = 1$ and find it is in good agreement with Prob(x). We find that $\mathrm{Im}\langle\pi_x\rangle$ is ten times smaller, making $\langle\pi_x\rangle$ largely real, as expected. Error bars are ±1 s.d. found from statistics in repeated scans. In **b**, only every third error bar is shown for clarity.

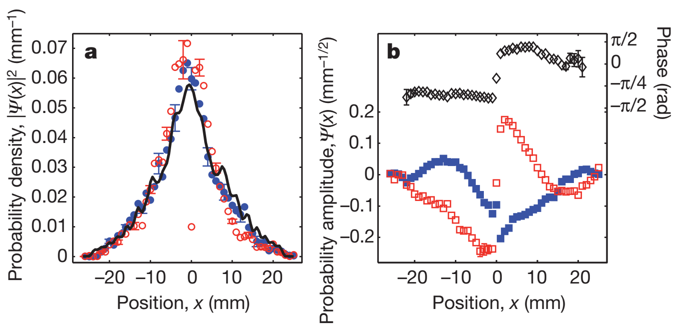


Figure 3 | Measurements of modified wavefunctions. We further test our ability to measure $\Psi(x)$ by changing Prob(x) by placing a reverse bull's-eye spatially apodized attenuator (RB in Fig. 1) (Edmund Optics, NT64-388) after the fibre. **a**, We calculate $|\Psi(x)|^2$ from the data (solid blue circles) along with a detector scan of Prob(x) (solid line) and find good agreement between the two. **b**, With the reverse bull's eye still in place, we modify the phase profile $\phi(x)$ of the wavefunction by creating a phase discontinuity at $x = 0$ imposed with a glass plate half-way across $\Psi(x)$. At the bottom, we show Re$\Psi(x)$ (solid blue squares; left axis) and Im$\Psi(x)$ (open red squares; left axis), which exhibit a discontinuity at the plate edge. This discontinuity is even clearer in the phase difference between the wavefunctions measured with and without the glass plate, shown at the top (open black diamonds; right axis). Despite their discontinuities, if we use Re$\Psi(x)$ and Im$\Psi(x)$ to calculate $|\Psi(x)|^2$ (open red circles in **a**), we find that it is largely unchanged by the glass plate. This is as expected, as the glass has a transmission near unity. Error bars are ± 1 s.d. found from statistics in repeated scans. In **a**, every third error bar is shown for clarity. In **b**, those bars smaller than the symbols are not shown.

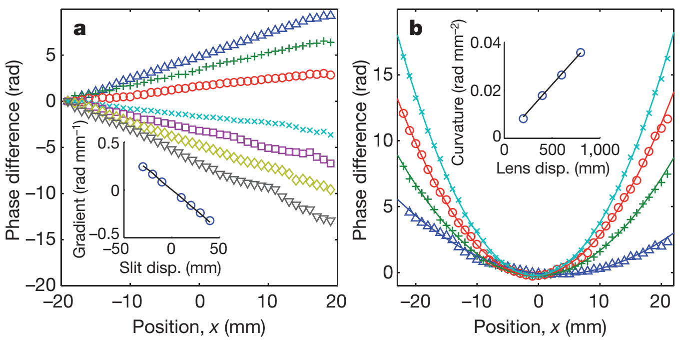


Figure 4 | Phase modification of the wavefunction. **a**, We displace the slit transversely by $\Delta x_{\text{slit}} = -30, -20, -10, 10, 20, 30$ and $40\,\mu$m (curves in main panel, top to bottom). This effectively redefines the zero momentum axis of the system. Our photons now travel at an angle to this axis or equivalently the wavefunction has a linear phase gradient, $\phi(x) = mx$, where $m = \Delta x_{\text{slit}} 2\pi/f_2\lambda$. We plot the phase difference between the original wavefunction and those with a phase gradient. For clarity, the curves have been offset to cross at -20 mm. This corresponds to shifting the arbitrary global phase of $\Psi(x)$. Inset, gradient m as a function of Δx_{slit} (circles) along with theory (line), which show good agreement. **b**, We introduce a quadratic phase by displacing the first lens by $\Delta z = 200, 400, 600$ and $800\,\mu$m (curves in main panel, bottom to top) along with theoretical fits (lines). The phase $\phi(x) = rx^2$, where the phase curvature $r = \pi\Delta z/f_1^2\lambda$. Inset, the phase curvature r from these fits (circles) as a function of lens displacement, Δx_{slit}, which shows good agreement with theory (line). Statistical error bars (± 1 s.d.) are smaller than the symbols in all plots.

good agreement between the expected and measured wavefunction, including its phase and magnitude (see the figure legends for details).

We now describe how the technique of weak measurement can be used to directly measure the quantum state of an arbitrary quantum system. We have the freedom to measure the quantum state in any chosen basis $\{|a\rangle\}$ (associated with observable A) of the system. The method entails weakly measuring a projector in this basis, $\pi_a \equiv |a\rangle\langle a|$, and post-selecting on a particular value b_0 of the complementary observable B (see Supplementary Discussion for a precise definition of complementarity). In this case, the weak value is

$$\langle\pi_a\rangle_{\text{W}} = \frac{\langle b_0|a\rangle\,\langle a|\Psi\rangle}{\langle b_0|\Psi\rangle} = \langle a|\Psi\rangle/\nu \qquad (6)$$

where ν is a constant, independent of a. Thus the weak value is proportional to the amplitude of state $|a\rangle$ in the quantum state. Stepping a through all the states in the A basis directly gives the quantum state represented in that basis:

$$|\Psi\rangle = \nu \sum_a \langle\pi_a\rangle_{\text{W}}|a\rangle \qquad (7)$$

This is the general theoretical result of this Letter. It shows that in any physical system one can directly measure the quantum state of that system by scanning a weak measurement through a basis and appropriately post-selecting in the complementary basis.

Weak measurement necessarily trades efficiency for accuracy or precision. A comparison of our method to current tomographic reconstruction techniques will require careful consideration of the signal to noise ratio in a given system. In order to increase this ratio in the direct measurement of the photon spatial wavefunction, future experiments will investigate the simultaneous post-selection of many transverse momenta.

In our direct measurement method, the wavefunction manifests itself as shifts of the pointer of the measurement apparatus. In this sense, the method provides a simple and unambiguous operational definition[19] of the quantum state: it is the average result of a weak measurement of a variable followed by a strong measurement of the complementary variable. We anticipate that the simplicity of the method will make feasible the measurement of quantum systems (for example, atomic orbitals, molecular wavefunctions[29], ultrafast quantum wavepackets[30]) that previously could not be fully characterized. The method can also be viewed as a transcription of quantum state of the system to that of the pointer, a potentially useful protocol for quantum information.

Received 22 March; accepted 14 April 2011.

1. Cohen-Tannoudji, C., Diu, B. & Laloe, F. *Quantum Mechanics* Vol. 1, 19 (Wiley-Interscience, 2006).
2. Mermin, N. D. What's bad about this habit. *Phys. Today* **62,** 8–9 (2009).
3. Landau, L. D. & Lifshitz, E. M. *Course of Theoretical Physics* Vol. 3, *Quantum Mechanics: Non-Relativistic Theory* 3rd edn, 6 (Pergamon, 1989).
4. Vogel, K. & Risken, H. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys. Rev. A* **40,** 2847–2849 (1989).
5. Smithey, D. T., Beck, M., Raymer, M. G. & Faridani, A. Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: application to squeezed states and the vacuum. *Phys. Rev. Lett.* **70,** 1244–1247 (1993).
6. Breitenbach, G., Schiller, S. & Mlynek, J. Measurement of the quantum states of squeezed light. *Nature* **387,** 471–475 (1997).
7. White, A. G., James, D. F. V., Eberhard, P. H. & Kwiat, P. G. Nonmaximally entangled states: production, characterization, and utilization. *Phys. Rev. Lett.* **83,** 3103–3107 (1999).
8. Hofheinz, M. *et al.* Synthesizing arbitrary quantum states in a superconducting resonator. *Nature* **459,** 546–549 (2009).
9. Aharonov, Y., Albert, D. Z. & Vaidman, L. How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. *Phys. Rev. Lett.* **60,** 1351–1354 (1988).
10. Ritchie, N. W. M., Story, J. G. & Hulet, R. G. Realization of a measurement of a "weak value". *Phys. Rev. Lett.* **66,** 1107–1110 (1991).
11. Resch, K. J., Lundeen, J. S. & Steinberg, A. M. Experimental realization of the quantum box problem. *Phys. Lett. A* **324,** 125–131 (2004).
12. Smith, G. A., Chaudhury, S., Silberfarb, A., Deutsch, I. H. & Jessen, P. S. Continuous weak measurement and nonlinear dynamics in a cold spin ensemble. *Phys. Rev. Lett.* **93,** 163602 (2004).
13. Pryde, G. J., O'Brien, J. L., White, A. G., Ralph, T. C. & Wiseman, H. M. Measurement of quantum weak values of photon polarization. *Phys. Rev. Lett.* **94,** 220405 (2005).
14. Mir, R. *et al.* A double-slit 'which-way' experiment on the complementarity-uncertainty debate. *N. J. Phys.* **9,** 287 (2007).
15. Hosten, O. & Kwiat, P. Observation of the spin Hall effect of light via weak measurements. *Science* **319,** 787–790 (2008).
16. Dixon, P. B., Starling, D. J., Jordan, A. N. & Howell, J. C. Ultrasensitive beam deflection measurement via interferometric weak value amplification. *Phys. Rev. Lett.* **102,** 173601 (2009).
17. Lundeen, J. S. & Steinberg, A. M. Experimental joint weak measurement on a photon pair as a probe of Hardy's paradox. *Phys. Rev. Lett.* **102,** 020404 (2009).

18. Aharonov, Y., Popescu, S. & Tollaksen, J. A time-symmetric formulation of quantum mechanics. *Phys. Today* **63,** 27–32 (2010).
19. Bridgman, P. *The Logic of Modern Physics* (Macmillan, 1927).
20. Wootters, W. K. & Zurek, W. H. A single quantum cannot be cloned. *Nature* **299,** 802–803 (1982).
21. Trebino, R. *Frequency-Resolved Optical Gating: The Measurement of Ultrashort Laser Pulses* (Springer, 2002).
22. Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409,** 46–52 (2001).
23. Duan, L. M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414,** 413–418 (2001).
24. Aharonov, Y. & Vaidman, L. Properties of a quantum system during the time interval between two measurements. *Phys. Rev. A* **41,** 11–20 (1990).
25. Lundeen, J. S. & Resch, K. J. Practical measurement of joint weak values and their connection to the annihilation operator. *Phys. Lett. A* **334,** 337–344 (2005).
26. Jozsa, R. Complex weak values in quantum measurement. *Phys. Rev. A* **76,** 044103 (2007).
27. Mukamel, E., Banaszek, K., Walmsley, I. A. & Dorrer, C. Direct measurement of the spatial Wigner function with area-integrated detection. *Opt. Lett.* **28,** 1317–1319 (2003).
28. Smith, B. J., Killett, B., Raymer, M. G., Walmsley, I. A. & Banaszek, K. Measurement of the transverse spatial quantum state of light at the single-photon level. *Opt. Lett.* **30,** 3365–3367 (2005).
29. Itatani, J. *et al.* Tomographic imaging of molecular orbitals. *Nature* **432,** 867–871 (2004).
30. Dudovich, N. *et al.* Measuring and controlling the birth of attosecond XUV pulses. *Nature Phys.* **2,** 781–786 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Natural Sciences and Engineering Research Council and the Business Development Bank of Canada.

Author Contributions The concept and the theory were developed by J.S.L. All authors contributed to the design and building of the experiment and the text of the manuscript. J.S.L, B.S. and C.B. performed the measurements and the data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.S.L. (jeff.lundeen@nrc-cnrc.gc.ca).

doi:10.1038/nature10173

Hydrogen bonding at the water surface revealed by isotopic dilution spectroscopy

Igor V. Stiopkin[1]†, Champika Weeraman[1]†, Piotr A. Pieniazek[2], Fadel Y. Shalhout[1]†, James L. Skinner[2] & Alexander V. Benderskii[1]

The air–water interface is perhaps the most common liquid interface. It covers more than 70 per cent of the Earth's surface and strongly affects atmospheric, aerosol and environmental chemistry. The air–water interface has also attracted much interest as a model system that allows rigorous tests of theory, with one fundamental question being just how thin it is. Theoretical studies have suggested a surprisingly short 'healing length' of about 3 ångströms (1 Å = 0.1 nm), with the bulk-phase properties of water recovered within the top few monolayers[1–3]. However, direct experimental evidence has been elusive owing to the difficulty of depth-profiling the liquid surface on the ångström scale. Most physical, chemical and biological properties of water, such as viscosity, solvation, wetting and the hydrophobic effect, are determined by its hydrogen-bond network. This can be probed by observing the lineshape of the OH-stretch mode, the frequency shift of which is related to the hydrogen-bond strength[4–6]. Here we report a combined experimental and theoretical study of the air–water interface using surface-selective heterodyne-detected vibrational sum frequency spectroscopy to focus on the 'free OD' transition found only in the topmost water layer. By using deuterated water and isotopic dilution to reveal the vibrational coupling mechanism, we find that the free OD stretch is affected only by intramolecular coupling to the stretching of the other OD group on the same molecule. The other OD stretch frequency indicates the strength of one of the first hydrogen bonds encountered at the surface; this is the donor hydrogen bond of the water molecule straddling the interface, which we find to be only slightly weaker than bulk-phase water hydrogen bonds. We infer from this observation a remarkably fast onset of bulk-phase behaviour on crossing from the air into the water phase.

The vibrational spectroscopy of aqueous interfaces has progressed significantly in recent years with the development of surface-selective spectroscopic techniques[7–12], and with the ability to model the dynamic H-bond network of bulk-phase and interfacial water and improved methods of calculating spectroscopic signals[2,13–16]. These tools can be applied to the free OH-stretch mode at the air–water interface[17,18], which presents a unique probe of the surface hydrogen bonds because this relatively narrow transition is spectrally well separated from the much broader and redshifted hydrogen-bonded part of the OH-stretch band. The relatively weak coupling of the free OH to other OH-stretch modes permits frequency-domain spectroscopic analysis of its lineshape, as discussed below. Furthermore, water molecules with a free OH are only found in the topmost monolayer at the interface, eliminating the ambiguity of the probe location (depth) within the interfacial region.

The intramolecular and intermolecular vibrational coupling between the OH transition dipoles of the same molecule or between neighbours affect the spectral lineshapes of the water OH-stretch band[19,20]. To disentangle these vibrational coupling contributions to the lineshape, we measure the free OD spectra at the air–water interface while using isotopic dilution in D_2O:HOD: H_2O mixtures to gradually turn off the vibrational coupling (Fig. 1)[21,22]. This, however, challenges spectroscopic detection limits. The free OD (or free OH) spectral feature can only be observed using the surface-selective vibrational sum frequency generation (SFG) technique[17,18]. Because SFG is a coherent optical process, the signal intensity decreases quadratically with dilution. Moreover, at low concentration, SFG spectra are overwhelmed by the nonresonant background, which interferes with the resonant OD-stretch signal.

We overcome these challenges by using the heterodyne-detected SFG technique[10], which uses interference of the signal with a reference beam to (1) amplify the SFG signal, considerably enhancing the sensitivity, (2) make it linear with the number of chromophores, and (3) allow separation of the resonant part of the signal from the nonresonant background by providing both the amplitude and the phase (or, alternatively, real and imaginary parts) of the signal[8,10,12,22]. This approach allowed us to record background-free SFG spectra with a sufficiently high signal-to-noise ratio to reveal the structure of the free OD-stretch lineshapes at the air–water interface of isotopic mixtures. As detailed in the Supplementary Information, we used broad-band spectral interferometry[23] to record heterodyne-detected SFG signals[10], and a delay between the infrared and visible pulses that simultaneously improves the SFG signal and spectral resolution and suppresses the nonresonant background[24].

Our main finding is that two distinct peaks are observed in the free OD-stretch spectrum of the air–water interface of D_2O:HOD:H_2O isotopic mixtures; the relative amplitudes of these peaks change with isotopic dilution (Fig. 2) for both SSP and PPP polarization combinations ('S' and 'P' indicate polarization perpendicular and parallel to the plane of incidence, respectively, for the SFG, visible and infrared beams). We performed curve fitting of both the real and imaginary parts of the SFG signals to extract the peak positions and amplitudes. The resonant part of the nonlinear susceptibility χ^2 is described as a

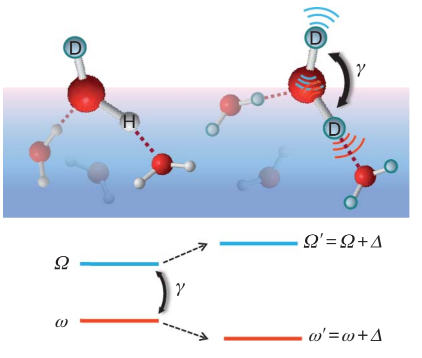


Figure 1 | The vibrational coupling scheme of the free OD stretch at the air–water interface. Ω is the free OD frequency, ω is the frequency of the coupling partner mode (the other OD stretch), and Δ is the frequency shift; γ is the coupling strength. D/H are indicated as blue/white spheres, O as red spheres. Dotted lines pictorially represent hydrogen bonds.

[1]Department of Chemistry, Wayne State University, Detroit, Michigan 48202, USA. [2]Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53705, USA. †Present addresses: Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53705, USA (I.V.S.); Steacie Institute for Molecular Sciences, National Research Council, Ottawa, Ontario, K1A OR6 Canada (C.W.); Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA (F.Y.S. and A.V.B.).

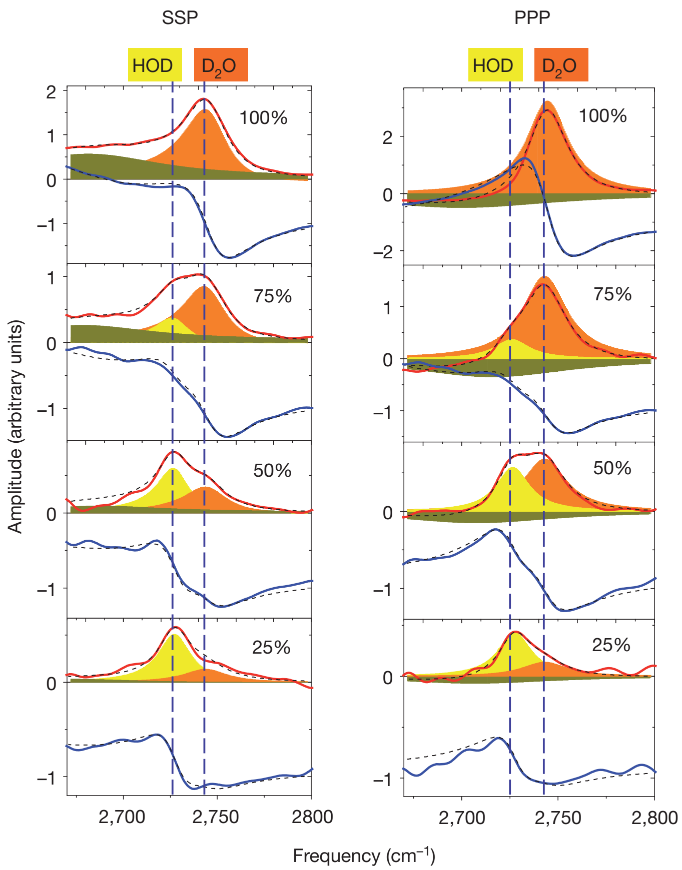


Figure 2 | The spectra of the free OD stretch at the air–water interface. The heterodyne-detected vibrational SFG spectra of the free OD stretch are shown for different isotopic mixtures H_2O:HOD:D_2O, with the D/H mole fraction indicated in each spectrum (as a percentage). The left panels show spectra recorded for the SSP polarization combination of SFG–visible–infrared beams; the right panels show PPP spectra. The solid blue and red lines are the experimentally measured real and imaginary parts of the SFG signal. Dashed lines represent the fits to a sum of Lorentzian terms and a nonresonant background:

$$\chi^{(2)}(\omega_{\mathrm{infrared}}) \propto A_{\mathrm{nonresonant}} e^{i\phi_{\mathrm{nonresonant}}} + \sum_j \frac{B_j \Gamma_j}{(\omega_{\mathrm{infrared}} - \Omega_j) + i\Gamma_j}$$

as described in the text. The components of the resonant part of the response used in the fitting are shown as shaded Lorentzian peaks. The two narrow peaks with interchanging amplitudes (shaded orange and yellow) are assigned, respectively, to the free OD of D_2O and HOD molecules at the interface, as detailed in the text.

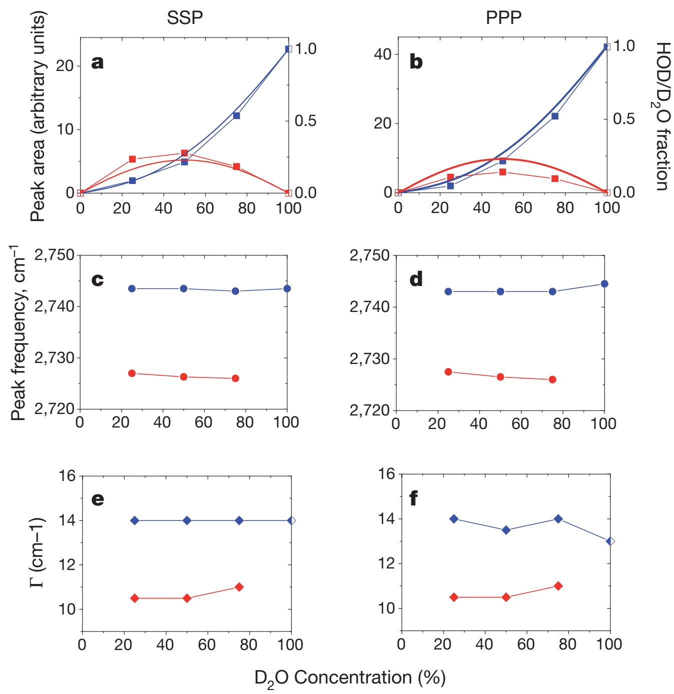


Figure 3 | Changes in vibrational spectra with isotopic dilution. **a**, **b**, Peak areas from spectral fitting (Lorentzian $B_j\Gamma_j$) of the free OD stretch of the DOD peak (blue squares), and the free OD stretch of the HOD peak (red squares), as a function of isotopic dilution. The scaling expected from isotopic scrambling, taking into account that D_2O has two potential free ODs while HOD has one, is shown as solid lines (right axis). **c**, **d**, Peak frequencies (Ω_j from Lorentzian fit) of the free OD stretch of the DOD peak (blue circles) and the free OD stretch of the HOD peak (red circles) as a function of isotopic dilution. **e**, **f**, Linewidths (Γ_j from Lorentzian fit) of the free OD of DOD peak (blue diamonds) and free OD of HOD peak (red diamonds) as a function of isotopic dilution. The left panels show the fitting results for SSP spectra and the right panels show the fitting results for PPP spectra. The lines between experimental points are to guide the eye only.

sum of Lorentzian terms, whereas the nonresonant background is represented as a constant term in the real part of the signal. Heterodyne-detected SFG enables us to separate the nonresonant background in the SFG spectra, which is crucial to revealing the true resonant lineshapes of the peak structure presented in Fig. 2.

The fitting results are summarized in Fig. 3. The observed peak amplitudes track the expected HOD:D_2O isotopic scrambling ratio (Fig. 3a, b), taking into account that D_2O molecules have two OD stretches that can be exhibited at the surfaces as free OD whereas HOD molecules have one. We therefore assign the $2{,}745 \pm 2\,cm^{-1}$ peak observed in pure D_2O to the free OD stretch of the D_2O molecule, and the $2{,}728 \pm 2\,cm^{-1}$ peak to the free OD stretch of the HOD molecule. The amplitude of the HOD peak first increases and then decreases with dilution, as expected.

The free OD frequency of HOD molecules, $\Omega_{\mathrm{HOD}} = 2{,}728\,cm^{-1}$, coincides within experimental uncertainty with the OD-stretch local mode of $2{,}727\,cm^{-1}$ of HOD molecules in the gas phase[25]. This suggests that the free OD mode of HOD at the air–water interface is essentially decoupled from all other vibrational modes in the system. Indeed, in a dilute HOD:H_2O mixture, the free OD is far off-resonance with the neighbouring OH-stretch vibrational modes (about $1{,}000\,cm^{-1}$ higher in frequency).

The free OD frequency of D_2O molecules, $\Omega_{\mathrm{DOD}} = 2{,}745\,cm^{-1}$, is blueshifted by $\Delta = 17 \pm 1.5\,cm^{-1}$ from that of HOD. The sign of the shift can be understood in terms of a simple model of two coupled oscillators (Fig. 1). Because the free OD mode Ω_{OD} occurs on the extreme blue side of the OD-stretch band, the majority of possible coupling partners—that is, the nearby OD-stretch modes with the frequency ω_{OD}—are likely to be of lower frequency. In the absence of coupling (free OD of HOD diluted in H_2O), the two modes are described by a (2×2) Hamiltonian matrix:

$$H_0 = \begin{pmatrix} \Omega & 0 \\ 0 & \omega \end{pmatrix}, \Omega > \omega$$

Switching on the coupling

$$H' = \begin{pmatrix} 0 & \gamma \\ \gamma & 0 \end{pmatrix}$$

leads to the blueshift of the higher-frequency mode and an equal redshift of the lower-frequency mode, so that $\Omega' = \Omega + \Delta$ and $\omega' = \omega - \Delta$, with

$$\Delta = \sqrt{\left(\frac{\delta}{2}\right)^2 + \gamma^2} - \frac{\delta}{2} \approx \frac{\gamma^2}{\delta} (\text{when } \gamma \ll \delta) \quad (1)$$

where $\delta = \Omega - \omega$ is the frequency mismatch between the two coupled modes. We note that Δ is positive regardless of the sign of the coupling

strength γ, consistent with the blueshift of the free OD of D_2O (the coupled case) with respect to the completely decoupled free OD of HOD.

The free OD frequency of DOD can be coupled either to OD stretches on other water molecules owing to intermolecular (predominantly dipole–dipole) interactions or to the other OD stretch on the same D_2O molecule. Although their amplitudes change, the peak frequencies Ω_{HOD} and Ω_{DOD} do not exhibit appreciable shifts on isotopic dilution, within the experimental uncertainty (Fig. 3c and d). This immediately rules out intermolecular vibrational coupling as the source of the $\Delta = 17\,cm^{-1}$ shift. Indeed, in that case, the coupling strength γ and the amount of the shift would depend on the average distance between the OD chromophores and would change by nearly an order of magnitude as concentration decreases from 100% D_2O to 25% D_2O:75% H_2O (corresponding to D_2O: HOD:$H_2O \approx$ 1:6:9).

We therefore conclude that the $\Delta = 17\,cm^{-1}$ shift between the free OD stretch of HOD and of D_2O at the air–water interface is caused by the intramolecular coupling of the free OD stretch to the other OD stretch on the same D_2O molecule. Through this intramolecular coupling, the free OD stretch reflects the frequency ω' of the other OD stretch (equation (1)) and hence the strength of the donor hydrogen bond of the D_2O molecule straddling the interface (Fig. 1).

We performed molecular dynamics simulations and theoretical modelling of the spectra to analyse the experimental results further and to provide a molecular view of the vibrational excitations and hydrogen bonds at the interface. Spectral simulations used the mixed quantum/classical approach and local electric field maps for spectroscopic parameters such as frequency, transition dipole and coupling, as described previously[13,26]. (Details are provided in the Supplementary Information.) The following criteria were used to identify the molecules containing the free OD oscillator: (1) the OD-stretch frequency is above 2,680 cm^{-1}; (2) the molecule is positioned within 6 Å of the Gibbs dividing surface; (3) the angle between the free OD vector and the vector from D to the nearest oxygen of another water molecule is larger than 90°. Analysis of the molecular dynamics trajectories shows that 23% of water molecules at the air–water interface have free OD (Fig. 4a), in agreement with earlier molecular dynamics simulations[2] and experimental estimates[17].

The calculated spectra reproduce the experimentally observed trends in a nearly quantitative fashion, as illustrated by the comparison between the calculated SSP spectra for pure D_2O and dilute HOD in H_2O and the experimental spectra of 100% D_2O and a 25% D_2O: 75% H_2O mixture (Fig. 4b and c). The simulated spectra show a blueshift of about 12 cm^{-1} and additional broadening (about 3 cm^{-1}) of the free OD stretch line of D_2O versus the uncoupled case of HOD in H_2O. The experimental Lorentzian linewidth of the free OD of HOD is $\Gamma_{HOD} = 11 \pm 1.5\,cm^{-1}$, whereas for D_2O it is $\Gamma_{DOD} = 14 \pm 1\,cm^{-1}$ (Fig. 3e and f). The theoretical linewidths are broader than the experimental ones, probably because the map of frequency versus local electric field used to calculate the spectra is imperfect.

The mixed quantum/classical methodology allows us to turn on/off selectively the intermolecular and intramolecular vibrational coupling and observe their respective effects on the calculated lineshapes. The spectral densities shown in Fig. 4d (lifetime effects and motional narrowing not included) identify the intramolecular coupling between the free OD bond and the other OD bonds of the same molecule as mainly responsible for the frequency shift and broadening of the free OD stretch of pure D_2O versus HOD in H_2O. Most of the shift (about 9 cm^{-1} of the total calculated 12 cm^{-1} shift) is reproduced by including intramolecular coupling only, whereas the shift of 2 cm^{-1} when retaining only intermolecular coupling is insufficient to account for the experimental observation.

The distribution of the strengths of vibrational couplings involving OD in Fig. 4e shows that the intermolecular (predominantly dipole–dipole) coupling for the free OD oscillators peaks at around 5 cm^{-1} (in contrast, bulk-phase water molecules average about 18 cm^{-1}). That

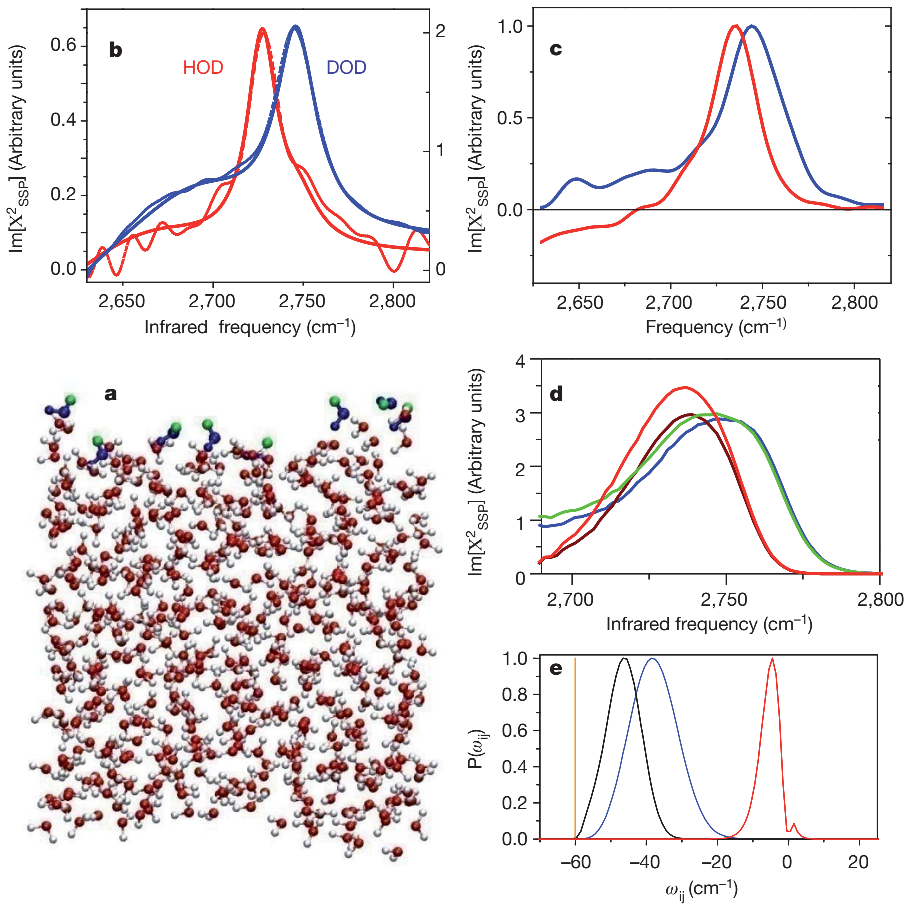


Figure 4 | Simulating and modelling the air–water interface. **a**, A snapshot of the molecular dynamics trajectory showing the air–water interface and molecules identified as having free OD bonds according to the definition in the text (blue); free OD bonds are shown in green. **b**, Experimentally recorded Im[χ^2] SSP spectra of the air–water interface of pure D_2O (blue) and a 25% D_2O: 75% H_2O mixture (red). **c**, Theoretically calculated Im[χ^2] SSP spectra for the free OD stretch of the D_2O molecule in pure D_2O (blue) and the isolated surface HOD molecule in H_2O (red). **d**, Spectral density of the free OD stretch calculated for the uncoupled case of HOD in H_2O (red), the fully coupled case of DOD in D_2O (blue), and when including only intermolecular coupling (brown), and when including only intramolecular coupling (green). **e**, Distribution of the coupling strengths. Shown are the intramolecular coupling γ_S in surface D_2O molecules that have a free OD, that is, coupling between the free OD and the other OD (black); intramolecular coupling between OD stretches of bulk-phase water molecules (blue); intermolecular coupling of the free OD to OD stretches on other water molecules (red); and intramolecular coupling in the gas-phase D_2O molecule, $-60\,cm^{-1}$ (yellow).

the free OD stretch is not effectively coupled to the OD stretches on other water molecules can be understood in terms of its frequency mismatch with most of the hydrogen-bonded (redshifted) OD partners and its having fewer immediate neighbours at the interface compared to bulk-phase water. Also, the free OD transition dipole is about three times weaker than the average OD stretch in bulk-phase water owing to non-Condon effects[14,27,28]. The orientation of the free OD (orthogonal to the interface, on average) relative to potential coupling partners (OD stretches in the underlying layer of water molecules tend to be oriented in-plane[2]) probably also diminishes the intermolecular coupling. Free OD–free OD coupling is negligible because of the small transition dipole of such vibrations (which appears squared in the expression of the coupling in this case) and because of the large average distance between the free ODs: even in pure D_2O, the 23% surface coverage translates into an average separation of about 7 Å.

The simulations also reproduce the broad shoulder with positive imaginary part at about 2,680 cm^{-1} in the SSP spectrum of pure D_2O (Fig. 4b and c). This feature is suppressed in the uncoupled case of dilute HOD in H_2O, and is not observed in the PPP spectra (Fig. 2). As explained in the Supplementary Information, we tentatively assign this feature to D_2O molecules with two donor hydrogen bonds and one acceptor hydrogen bond. While such molecules are oriented with both hydrogens down (on average), the intramolecular coupling switches on an antisymmetric linear combination of the two local modes; this combination has a higher frequency and a transition dipole pointing up, giving a positive contribution to the imaginary part of χ^2.

Using an electrostatic map of the intramolecular coupling in D_2O obtained from *ab initio* calculations on water clusters[14] and the molecular dynamics simulations, we determine the intramolecular coupling between the free OD and the other OD stretch on the same D_2O molecule at the interface to be about $\gamma_S \approx -48\ cm^{-1}$ (Fig. 4e); this coupling is weaker than the gas-phase value of $\gamma \approx -60\ cm^{-1}$ (refs 25 and 29) owing to the local interfacial environment. Using the surface value for γ_S and equation (1), we estimate that the other OD stretch frequency is $\omega' \approx 2{,}580\ cm^{-1}$—putting it well within the OD stretch band of bulk-phase water, which is centred at 2,480 cm^{-1} and broadened by hydrogen bonding to the full-width at half-maximum (FWHM) of about 300 cm^{-1} (ref. 30). This finding implies that one of the first hydrogen bonds encountered in the topmost layer at the water surface is only slightly weaker than the bulk-phase average. We note that although the other OD stretch is probably broadened by hydrogen bonding, this does not lead to significant broadening of the free OD: for example, according to equation (1), a width of about 50 cm^{-1} of the other OD band would result in only about 5 cm^{-1} broadening of the free OD bond. This is the probable mechanism for the additional width of the free OD of D_2O versus that of HOD (Fig. 3e and f).

We also note that for the stretching mode of free OH around 3,700 cm^{-1}, the vibrational coupling model presented here would predict a smaller shift ($\Delta \approx 10\ cm^{-1}$) between this mode in HOD and in H_2O, compared to $\Delta = 17\ cm^{-1}$ for the free OD, owing to (1) weaker intramolecular coupling γ in H_2O versus D_2O (about $-50\ cm^{-1}$ versus $-60\ cm^{-1}$ in the gas phase)[25,29], and (2) a larger frequency mismatch $\delta = \Omega - \omega$ between the free OH and the hydrogen-bonded region of the OH-stretch band (about 300 cm^{-1} in H_2O) versus about 250 cm^{-1} in D_2O. Taking into account the width of the free OH feature, this prediction would make it more difficult to observe two distinct peaks for the free OH of HOD and H_2O, consistent with previous reports[21,22].

Received 9 November 2010; accepted 28 April 2011.

1. Townsend, R. M., Gryko, J. & Rice, S. A. Structure of the liquid vapor interface of water. *J. Chem. Phys.* **82,** 4391–4392 (1985).
2. Morita, A. & Hynes, J. T. A theoretical analysis of the sum frequency generation spectrum of the water surface. *Chem. Phys.* **258,** 371–390 (2000).
3. Taylor, R. S., Dang, L. X. & Garrett, B. C. Molecular dynamics simulations of the liquid/vapor interface of SPC/E water. *J. Phys. Chem.* **100,** 11720–11725 (1996).
4. Rey, R., Moller, K. B. & Hynes, J. T. Hydrogen bond dynamics in water and ultrafast infrared spectroscopy. *J. Phys. Chem. A* **106,** 11993–11996 (2002).
5. Lawrence, C. P. & Skinner, J. L. Ultrafast infrared spectroscopy probes hydrogen-bonding dynamics in liquid water. *Chem. Phys. Lett.* **369,** 472–477 (2003).
6. Fecko, C. J., Eaves, J. D., Loparo, J. J., Tokmakoff, A. & Geissler, P. L. Ultrafast hydrogen-bond dynamics in the infrared spectroscopy of water. *Science* **301,** 1698–1702 (2003).
7. Richmond, G. L. Molecular bonding and interactions at aqueous surfaces as probed by vibrational sum frequency spectroscopy. *Chem. Rev.* **102,** 2693–2724 (2002).
8. Ostroverkhov, V., Waychunas, G. A. & Shen, Y. R. New information on water interfacial structure revealed by phase-sensitive surface spectroscopy. *Phys. Rev. Lett.* **94,** 046102 (2005).
9. Sovago, M., Campen, R. K., Bakker, H. J. & Bonn, M. Hydrogen bonding strength of interfacial water determined with surface sum-frequency generation. *Chem. Phys. Lett.* **470,** 7–12 (2009).
10. Stiopkin, I. V., Jayathilake, H. D., Bordenyuk, A. N. & Benderskii, A. V. Heterodyne-detected vibrational sum frequency generation spectroscopy. *J. Am. Chem. Soc.* **130,** 2271–2275 (2008).
11. Gan, W., Wu, D., Zhang, Z., Feng, R. R. & Wang, H. F. Polarization and experimental configuration analyses of sum frequency generation vibrational spectra, structure, and orientational motion of the air/water interface. *J. Chem. Phys.* **124,** 114705 (2006).
12. Nihonyanagi, S., Yamaguchi, S. & Tahara, T. Direct evidence for orientational flip-flop of water molecules at charged interfaces: a heterodyne-detected vibrational sum frequency generation study. *J. Chem. Phys.* **130,** 204704 (2009).
13. Auer, B. M. & Skinner, J. L. Vibrational sum-frequency spectroscopy of the liquid/vapor interface for dilute HOD in D_2O. *J. Chem. Phys.* **129,** 214705 (2008).
14. Corcelli, S. A., Lawrence, C. P. & Skinner, J. L. Combined electronic structure/molecular dynamics approach for ultrafast infrared spectroscopy of dilute HOD in liquid H_2O and D_2O. *J. Chem. Phys.* **120,** 8107–8117 (2004).
15. Nagata, Y. & Mukamel, S. Vibrational sum-frequency generation spectroscopy at the water/lipid interface: molecular dynamics simulation study. *J. Am. Chem. Soc.* **132,** 6434–6442 (2010).
16. Ishiyama, T. & Morita, A. Vibrational spectroscopic response of intermolecular orientational correlation at the water surface. *J. Phys. Chem. C* **113,** 16299–16302 (2009).
17. Du, Q., Superfine, R., Freysz, E. & Shen, Y. R. Vibrational spectroscopy of water at the vapor water interface. *Phys. Rev. Lett.* **70,** 2313–2316 (1993).
18. Du, Q., Freysz, E. & Shen, Y. R. Surface vibrational spectroscopic studies of hydrogen-bonding and hydrophobicity. *Science* **264,** 826–828 (1994).
19. Skinner, J. L., Auer, B. M. & Lin, Y. S. Vibrational line shapes, spectral diffusion, and hydrogen bonding in liquid water. *Adv. Chem. Phys.* **142,** 59–103 (2009).
20. Woutersen, S. & Bakker, H. J. Resonant intermolecular transfer of vibrational energy in liquid water. *Nature* **402,** 507–509 (1999).
21. Raymond, E. A., Tarbuck, T. L., Brown, M. G. & Richmond, G. L. Hydrogen-bonding interactions at the vapor/water interface investigated by vibrational sum-frequency spectroscopy of HOD/H_2O/D_2O mixtures and molecular dynamics simulations. *J. Phys. Chem. B* **107,** 546–556 (2003).
22. Tian, C. S. & Shen, Y. R. Isotopic dilution study of the water/vapor interface by phase-sensitive sum-frequency vibrational spectroscopy. *J. Am. Chem. Soc.* **131,** 2790–2791 (2009).
23. Lepetit, L., Cheriaux, G. & Joffre, M. Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy. *J. Opt. Soc. Am. B* **12,** 2467–2474 (1995).
24. Stiopkin, I. V., Jayathilake, H. D., Weeraman, C. & Benderskii, A. V. Temporal effects on spectroscopic line shapes, resolution, and sensitivity of the broad-band sum frequency generation. *J. Chem. Phys.* **132,** 234503 (2010).
25. Shimanouchi, T. *Tables of Molecular Vibrational Frequencies Consolidated* Vol. I (National Bureau of Standards, 1972).
26. Auer, B. M. & Skinner, J. L. Vibrational sum-frequency spectroscopy of the water liquid/vapor interface. *J. Phys. Chem. B* **113,** 4125–4130 (2009).
27. Corcelli, S. A. & Skinner, J. L. Infrared and Raman line shapes of dilute HOD in liquid H_2O and D_2O from 10 to 90 degrees C. *J. Phys. Chem. A* **109,** 6154–6165 (2005).
28. Loparo, J. J., Roberts, S. T., Nicodemus, R. A. & Tokmakoff, A. Variation of the transition dipole moment across the OH stretching band of water. *Chem. Phys.* **341,** 218–229 (2007).
29. Barker, E. F. & Sleator, W. W. The infrared spectrum of heavy water. *J. Chem. Phys.* **3,** 660–663 (1935).
30. Max, J. J. & Chapados, C. Isotope effects in liquid water by infrared spectroscopy. *J. Chem. Phys.* **116,** 4626–4642 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The experiments presented here were supported by the NSF CAREER grant number CHE-0449720 (I.V.S., C.W., F.Y.S. and A.V.B.). J.L.S. thanks the NSF and the DOE for support from grants CHE-0750307 and DE-FG02-09ER16110, respectively.

Author Contributions I.V.S., C.W., F.Y.S. and A.V.B. constructed the spectroscopic set-up, executed the experiments, and performed the analysis of the spectroscopic data. P.A.P. and J.L.S. carried out molecular dynamics simulations and calculation of the spectra. All authors discussed the results and contributed to the preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.V.B. (alex.benderskii@usc.edu).

LETTER

doi:10.1038/nature10150

Irregular tropical glacier retreat over the Holocene epoch driven by progressive warming

Vincent Jomelli[1], Myriam Khodri[2], Vincent Favier[3], Daniel Brunstein[1], Marie-Pierre Ledru[4], Patrick Wagnon[5], Pierre-Henri Blard[6], Jean-Emmanuel Sicart[5], Régis Braucher[7], Delphine Grancher[1], Didier Louis Bourlès[7], Pascale Braconnot[8] & Mathias Vuille[9]

The causes and timing of tropical glacier fluctuations during the Holocene epoch (10,000 years ago to present) are poorly understood. Yet constraining their sensitivity to changes in climate[1] is important, as these glaciers are both sensitive indicators of climate change and serve as water reservoirs for highland regions[2]. Studies have so far documented extra-tropical glacier fluctuations[3,4], but in the tropics, glacier–climate relationships are insufficiently understood. Here we present a ^{10}Be chronology for the past 11,000 years (11 kyr), using 57 moraines from the Bolivian Telata glacier (in the Cordillera Real mountain range). This chronology indicates that Telata glacier retreated irregularly. A rapid and strong melting from the maximum extent occurred from 10.8 ± 0.9 to 8.5 ± 0.4 kyr ago, followed by a slower retreat until the Little Ice Age, about 200 years ago. A dramatic increase in the rate of retreat occurred over the twentieth century. A glacier–climate model indicates that, relative to modern climate, annual mean temperature for the Telata glacier region was −3.3 ± 0.8 °C cooler at 11 kyr ago and remained −2.1 ± 0.8 °C cooler until the end of the Little Ice Age. We suggest that long-term warming of the eastern tropical Pacific and increased atmospheric temperature in response to enhanced austral summer insolation were the main drivers for the long-term Holocene retreat of glaciers in the southern tropics.

Accurate documentation of past glacier fluctuations in the tropics is of primary importance in understanding the climate processes and thresholds responsible for their mass balance variations. Recent papers[5–11] documenting Holocene deglaciation in the Andes on the basis of dated moraines and proglacial lake sediments show that glaciers were larger than today in the early and late Holocene. Even though these

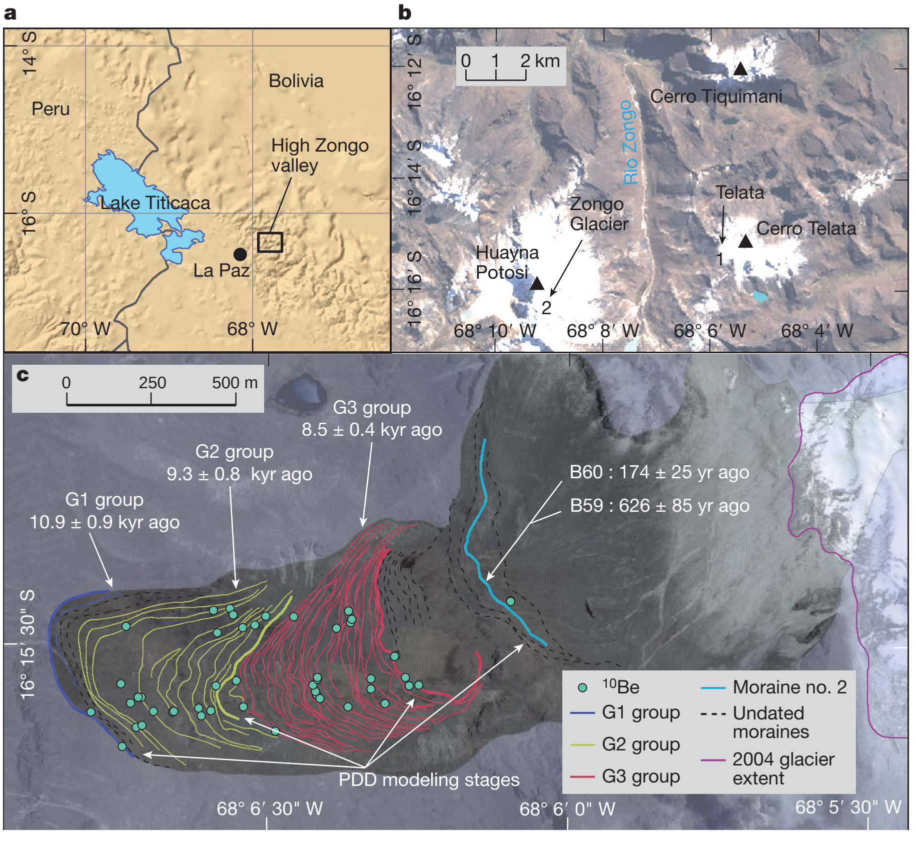


Figure 1 | Telata glacier and sampling sites. **a**, Location of the high Zongo valley. **b**, Local map showing the study sites: 1, the Telata glacier; 2, the location of the Zongo glacier; filled triangles, summits. **c**, Map of the Telata glacier, showing dated and undated Holocene moraines (see Supplementary Information sections 1 and 2, Supplementary Tables 1–5), the location of ^{10}Be sampling sites (green dots), the extent of the Telata glacier in 2004 and the moraines used to run the PDD model. Uncertainty associated with the age of each group accounts for analytical uncertainties only.

[1]Université Paris 1 Panthéon-Sorbonne-CNRS Laboratoire de Géographie Physique, 92195 Meudon, France. [2]IRD/IPSL/Laboratoire d'Océanographie et du Climat : Expérimentations et Approches Numériques (LOCEAN), 4, place Jussieu, 75252 Paris Cedex 05, France. [3]UJF-CNRS, Laboratoire de Glaciologie et de Géophysique de l'Environnement (LGGE), BP 96, 38402 St Martin d'Hères Cedex, France. [4]IRD UMR 226, Institut des Sciences de l'Evolution de Montpellier (ISEM), UM2 CNRS IRD, Place Eugène Bataillon cc061, 34095 Montpellier Cedex, France. [5]IRD/UJF-Grenoble 1/CNRS/G-INP, LGGE UMR 5183, LTHE UMR 5564, Grenoble, F-38402, France. [6]Centre de Recherches Pétrographiques et Géochimiques (CRPG), UPR2300, CNRS, Université de Lorraine, 54501 Vandoeuvre-lès-Nancy, France. [7]CEREGE/CNRS, Aix Marseille Université, UMR 6635, Europôle Méditerranéen de l'Arbois, Boîte Postale 80, 13545 Aix-en-Provence Cedex, France. [8]IPSL/Laboratoire des Sciences du Climat et de l'Environnement, Unité mixte CEA-CNRS-UVSQ, Orme des Merisiers, Bât. 709, CEA-SACLAY, 91191 Gif-sur-Yvette Cedex, France. [9]Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, New York 12222, USA.

studies give precious information, little attention has been paid so far to reconstructing glaciological parameters, such as glacier mass balance, which provide the direct link to regional- and global-scale climatic conditions and which allow the constraining of temperature/precipitation changes responsible for the observed glacier fluctuations over the past 11 kyr.

In Peru and Bolivia, several glaciers exhibit numerous moraines younger than the Last Glacial Maximum (Supplementary Table 1). Here, relying on a very well preserved and complete sequence of 57 successive moraines from the Telata glacier in the Bolivian Cordillera Real, we explore the issue of tropical glacier evolution and how glaciers responded to climatic changes during the Holocene (Fig. 1).

To allow a precise reconstruction of the palaeoglacier extents, we carefully mapped these 57 Telata moraines. Their chronology was established from ^{10}Be exposure age dating on 59 sampled boulders (Fig. 1; Supplementary Information section 1, Supplementary Figs 1–5, Supplementary Tables 1–5). In order to get enough replicates to better constrain the uncertainties, moraines were pooled into three groups (G1–G3; see Supplementary Information section 1). Analytical uncertainties on the entire set of ^{10}Be ages averaged 5.4 ± 5.7%.

The first group of moraines (G1) dated to 10.8 ± 0.9 kyr ago on the ^{10}Be chronology (that is, 10.8 ± 0.9 ^{10}Be kyr ago; two samples). These correspond to the late glacial outer moraine M57 with frontal termination located at 4,400 m above sea level (a.s.l.) and three younger ridges, all formed during retreat periods. Two samples collected on M57 and on M56 were viewed as outliers—on the basis of a χ^2 test reflecting post-depositional processes, and of isotope inheritance from previous exposure (Supplementary Figs 3 and 4, Supplementary Table 4)—and were therefore rejected. The second group of moraines (G2, M53–M40) was dated to 9.3 ± 0.8 ^{10}Be kyr ago on the basis of 24 ^{10}Be samples. Two outliers located on M53, and three others located on M51, M46 and M42, were rejected (post-depositional processes and inheritance from previous exposure). The third group (G3, M39–M11) was dated to around 8.5 ± 0.4 ^{10}Be kyr ago, on the basis of 21 ^{10}Be samples. One outlier on M39 and two others located on M26 and on M22 were rejected from the analysis (Supplementary Figs 3 and 4, Supplementary Table 4). These early Holocene moraine advances correlate well with ^{10}Be-dated moraines in the southern Andes[5–8], revealing a regional synchrony. The last two moraines were deposited about 626 ± 85 and 174 ± 25 years ago, synchronously with the Little Ice Age (LIA) moraines observed in the nearby Zongo valley[12,13]. However, to compare these mean ages with other chronological records, an additional 9% uncertainty (linked to the uncertainty on the production rate) has to be considered and added to the analytical uncertainties (Supplementary Information section 1).

Between M57 and the front created in 2004, the glacier retreated by about 2.9 km. This retreat was however irregular and occurred in three distinct steps. The distance between the late glacial (11 kyr ago) and the early Holocene (8.5 kyr ago) frontal moraines is about 1.1 km, which equates to a minimum loss of 37% of the total length during a relatively short period. By contrast, the distance between the 8.5-kyr-ago and LIA frontal moraines is about 0.4 km, corresponding to a minimum loss of 13% of the total length over a longer time interval covering most of the Holocene. Finally, in the recent past, since the last moraine was deposited about 200 years ago, the glacier has undergone a reduction of more than 30% of its initial length. Today, the Telata is about 500 m long with a total surface area of 0.137 km^2, compared to 0.89 km^2 200 years ago and 1.79 km^2 11 kyr before present.

These three phases of glacier retreat since 11 kyr ago may have been caused by different combinations of warming and/or decline in precipitation. To explore potential combinations of precipitation and temperature changes (and resulting local climatic conditions) that are consistent with the extents of the Telata glacier, we applied an improved positive degree day (PDD) model[14] (see Methods) calibrated with measurements made on the nearby Zongo glacier (Fig. 1). We applied this model for three distinct climatic periods: (1) during deglaciation, 11 kyr ago, (2) the early Holocene (9.3 and 8.5 kyr ago), and (3) during the LIA (Supplementary Information sections 1 and 2, Supplementary Table 7). This model integrates the effects of solar radiation, accumulation and albedo through changes in glacier surface characteristics[14]. Indeed, energy balance measurements in Bolivia show that albedo is the main controlling factor for glacier ablation[15–17]. Although the model treats several physical processes in a simplified way, analyses in the Zongo watershed show that the model yields accurate results at annual and longer timescales[14]. In order to precisely reconstruct the palaeopositions of the equilibrium-line altitude (ELA), the modelled mass balance law was projected on to the Telata topography by using an ice-flow model (Supplementary Text 1.3).

Using the present day climate data from Zongo, the model yields an ELA at 5,320 m a.s.l. (Fig. 2a), in good agreement with the currently observed ELA position of the Zongo glacier (Supplementary Information section 2). Model results for the late glacial moraine imply that 10.8 kyr ago, the ELA was at 4,800 m a.s.l, ~520 m below its current position. At 8.5 kyr ago and during the LIA (200 years ago), the ELA was located at 4,900 m a.s.l. and at 5,000 m a.s.l., respectively.

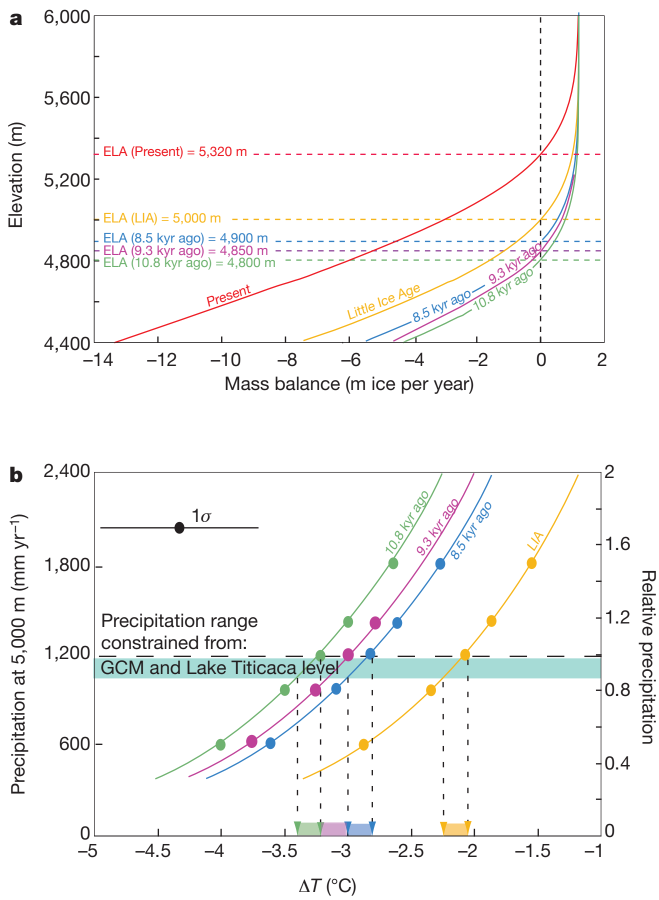


Figure 2 | Palaeoclimatic reconstructions of the Telata glacier. **a**, Plot of mass balance against elevation for the present day (red curve), 10.8 kyr ago (green), 9.3 kyr ago (purple), 8.5 kyr ago (blue) and Little Ice Age (yellow). Dashed lines, altitudes of the ELA from 10.8 kyr ago to present. The present-day mass balance curve at the Zongo glacier is characterized by an ELA at 5,320 m a.s.l. **b**, Palaeoclimatic conditions at 5,000 m a.s.l. required to reproduce the glacier extent during the past 10.8 kyr. Changes in temperature (ΔT) are plotted against changes in precipitation. Left-hand y axis, absolute precipitation at 5,000 m a.s.l.; right-hand y axis, relative precipitation at 5,000 m a.s.l., normalized to present day value. The 1σ error bar shows the temperature uncertainty arising from the model assumptions. Sensitivity tests indicate that the temperature uncertainty is ±0.7 °C. Dashed vertical lines with coloured arrows indicate the range of possible temperature depression for 10.8 kyr ago (green), 9.3 kyr ago (purple), 8.5 kyr ago (blue) and the Little Ice Age (yellow) under a scenario of 0–12% precipitation reduction.

The model does not, however, provide a unique solution for the inferred palaeoclimatic conditions, as there is a range of temperature/precipitation combinations that can maintain one given glacier extent. We thus used independent estimates of precipitation from six mid-Holocene PMIP-II general circulation model (GCM) simulations (Supplementary Information section 2, Supplementary Fig. 6). The GCM simulations show a 4–12% decrease in annual precipitation in the Cordillera Real compared to the pre-industrial period during the mid-Holocene. Additional simulations with one of the PMIP-II GCMs (IPSL-CM4[18]) for the early Holocene (9.5 kyr before present) reveals a 12% decrease, while no significant differences are simulated when comparing late Holocene (4 kyr before present) to pre-industrial conditions (not shown). Even though the amplitude varies, the consistency among PMIP-II models suggests that a slight precipitation decrease is a robust result for the mid-Holocene and might be independent of the GCM used (Supplementary Fig. 6). Moreover, previous studies depicted rather dry mid-Holocene conditions in the region compared to the late Holocene[11,19,20], in agreement with the results from the GCMs. Importantly, the relatively low levels of Lake Titicaca during the Holocene[10,20] indicate that the regional precipitation did not exceed its present level (Fig. 2)[21].

In order to assess the uncertainty arising from these palaeoprecipitation estimates, we used a range of precipitation changes varying from 88% to 100% of modern conditions for the whole Holocene period to constrain our glaciological model (Fig. 2). This approach indicates that within the range of uncertainties in precipitation changes, the glacier extents at 10.8 kyr ago, 8.5 kyr ago and during the LIA required a temperature change of $-3.3 \pm 0.8\,^{\circ}$C, $-2.9 \pm 0.8\,^{\circ}$C and $-2.1 \pm 0.8^{\circ}$C, respectively (Fig. 2b).

We observe that the tropical Telata glacier underwent a decline over the past 11 kyr that mirrors low-frequency insolation changes, suggesting a possible external forcing amplified by regional mechanisms (see below). The Telata glacier is indeed located in a region where precipitation is fed predominantly by summer monsoonal easterly airflow originating over the tropical Atlantic Ocean[22] (Fig. 3) (Supplementary Information section 2). The South American summer monsoon (SASM) is likely to have been the dominant moisture source for the glacier over the past 11 kyr (ref. 22), and glacier mass balance thus could have responded sensitively to precession-driven changes in SASM-related moisture supply. Documented regional climate changes attest that the early Holocene summer insolation minimum made the SASM weaker[18] than it was in the late Holocene period, favouring a northerly mean latitude for the intertropical convergence zone[23,24] (ITCZ; Fig. 4b, c) while cooling the eastern equatorial Pacific Ocean[24] (Fig. 4c). The rapid retreat of Telata between 11 and 8.5 kyr ago could have occurred in response to this quite rapid and significant decrease in SASM intensity. In addition to the summer insolation minimum, Northern Hemisphere land ice melting and the subsequent freshwater inflow in the North Atlantic from the late glacial to the early Holocene might have also favoured a southward shift of the ITCZ and the weakening of the SASM[23–25]. Coupled GCM simulations (Supplementary Information section 2) are able to reproduce the observed colder eastern equatorial Pacific Ocean and weaker SASM as a direct response to insolation changes only (Fig. 3).

On the other hand, our PDD model reveals that rather cold conditions over the Telata glacier in the late glacial, early Holocene and LIA were sufficient to maintain the observed past glacier extents. The recovering and strengthening of the SASM over the course of the Holocene period after 8.5 kyr before present must have led to a precipitation increase over the glacier. However, the moraine chronology shows that the glacier was still retreating during this period, despite increased moisture availability. This result suggests that the substantial increase in temperature and solar radiation from 8.5 kyr before present to the

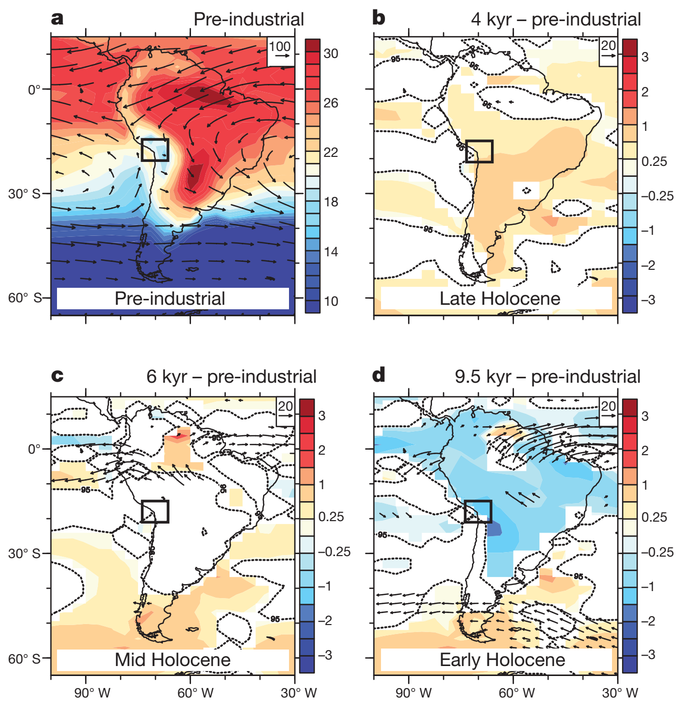


Figure 3 | Changes in surface temperature and atmospheric moisture transport simulated for December by the IPSL-CM4 GCM. The colour shading shows temperature in °C (**a**) or temperature difference (**b**–**d**). The dashed line indicates significant differences between Holocene and pre-industrial simulations at $P = 0.05$; vectors indicate the vertically integrated atmospheric latent heat transport in W m^{-2}. The reference vector is boxed at top right of each panel. **a**, Pre-industrial conditions; **b**, difference between 4 kyr BP and pre-industrial; **c**, difference between 6 kyr BP and pre-industrial; **d**, difference between 9.5 kyr BP and pre-industrial. Telata region shown boxed.

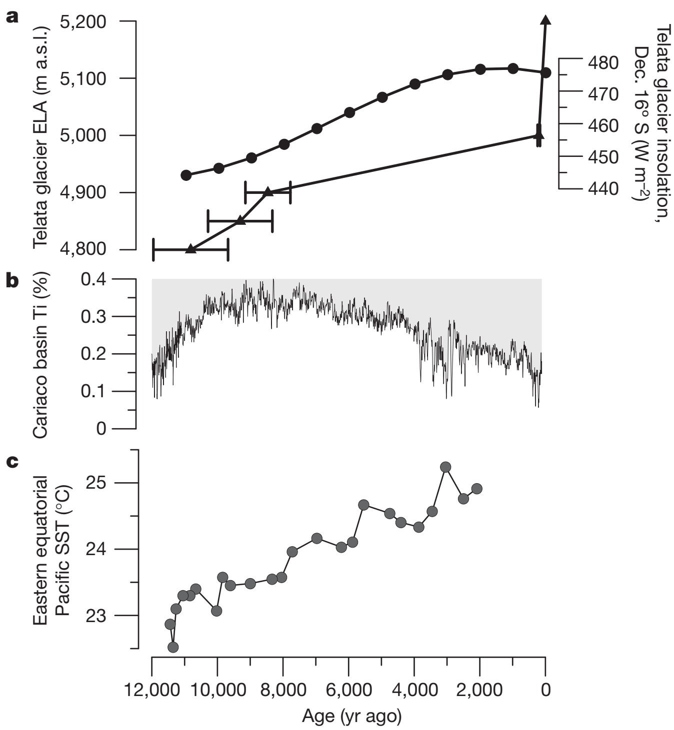


Figure 4 | Changes of the Telata glacier compared to tropical climate proxies. **a**, Triangles, changes in the ELA of the Telata glacier (frontal moraine dated by ^{10}Be); circles, December insolation changes at 16° S. Error bars (±s.e.m.) on each ^{10}Be group indicate uncertainty associated with the mean age and account for analytical uncertainty only. (Note that to these analytical uncertainties has to be added 9% uncertainty linked to the production rate when comparing theses ages with other chronological reconstructions.) **b**, Titanium concentration in Cariaco basin sediment[23]. **c**, SST in the eastern tropical Pacific[24].

late Holocene outweighed the impact of increased precipitation on the Telata glacier mass balance[26]. Energy balance measurements reveal that modern austral summer short-wave radiation controls glacier ablation and glacier mass balance through albedo feedbacks[15–17]. An increase in summer insolation would enhance regional temperature and affect the albedo (for example, through a shift in the rain–snow line) and hence increase glacier melt, in agreement with our PDD results. Although the lack of alternative proxy records precludes an independent confirmation of the temperature increase in the Telata region, the glacier retreat from the early Holocene to the LIA is consistent with progressively warmer eastern Pacific sea surface temperatures[24] (SSTs), a phenomenon also observed today[27] (Fig. 4a–c). The coupled GCM simulates an increase in surface temperature over both the eastern equatorial Pacific (+2 to +2.5 °C) and tropical South America (+1.5 to +2 °C) as a first order response to insolation forcing from the early Holocene to pre-industrial conditions (Figs 3, 4). A progressive warming throughout the Holocene is thus in agreement with the GCM results (Fig. 4).

Our results indicate that increasing insolation and a warming of 3.3 °C (±0.8 °C) was probably sufficient to counterbalance the strengthening of the SASM through the Holocene, thereby leading to the Telata glacier retreat. Future projections estimate a 4–5 °C warming in the high mountains of the tropical Andes[28,29] by 2100; this warming is larger than our estimate for the entire Holocene, which led to a rise in ELA of 520 m and a decrease of 92% in the surface area (a 2.4 km retreat) of the Telata glacier.

METHODS SUMMARY

Palaeoglacial extents were documented from moraine records. Ice volumes, mass balance variations with elevation, ELA and the associated climatic conditions corresponding to palaeomoraines were estimated with the PDD model combined with a dynamic ice flow model[14]. Current observations on the Zongo glacier reveal that variables affecting albedo directly, such as precipitation, or indirectly, such as temperature (through the elevation shift of the rain–snow line), have a strong effect on ablation and hence on the specific mass balance. In addition, current observations also demonstrate that austral summer precipitation is another important factor controlling accumulation over Bolivian glaciers at interannual timescales and therefore variability of the glacier's mass balance[27]. However, monthly and annual air temperature are significantly correlated with glacier mass balance because they are well correlated with the main energy fluxes through the ablation and accumulation processes over tropical glaciers at such timescales[17]. Hence, we used the PDD model that assumes both precipitation and temperature as fundamental variables of the mass balance. This model also integrates the effect of solar radiation and uses different melting factors for ice and snow[14]. For the period 1991–2008, the annual mean temperature measured on the lateral moraine of the Zongo glacier at 5,000 m a.s.l. was ~0.5 °C, and mean annual precipitation was 1,200 mm yr^{-1}.

Received 3 March; accepted 21 April 2011.

1. Vuille, M. *et al.* Climate change and tropical Andean glaciers: past, present and future. *Earth Sci. Rev.* **89,** 79–96 (2008).
2. Kaser, G. & Osmaston, H. *Tropical Glaciers* (Cambridge Univ. Press, 2002).
3. Mayewski, P. A. *et al.* Holocene climate variability. *Quat. Res.* **62,** 243–255 (2004).
4. Schaefer, J. M. *et al.* High-frequency Holocene glacier fluctuations in New Zealand differ from the northern signature. *Science* **324,** 622–625 (2009).
5. Smith, C. A., Lowell, T. V., Owen, L. A. & Caffee, M. W. Late Quaternary glacial chronology on Nevado Illimani, Bolivia, and the implications for paleoclimatic reconstructions across the Andes. *Quat. Res.* **75,** 1–10 (2011).
6. Glasser, N. F., Clemmens, S., Schnabel, C., Fenton, C. R. & McHargue, L. Tropical glacier fluctuations in the Cordillera Blanca, Peru between 12.5 and 7.6 ka from cosmogenic ^{10}Be dating. *Quat. Sci. Rev.* **28,** 3448–3458 (2009).
7. Licciardi, J. M., Schaefer, J. M., Taggart, J. R. & Lund, D. C. Holocene glacier fluctuations in the Peruvian Andes indicate northern climate linkages. *Science* **325,** 1677–1679 (2009).
8. Rodbell, D. T., Smith, J. A. & Mark, B. G. Glaciation in the Andes during the late glacial and Holocene. *Quat. Sci. Rev.* **28,** 2165–2212 (2009).
9. Buffen, A. M., Thompson, L. G., Mosley-Thompson, E. & Huh, K. I. Recently exposed vegetation reveals Holocene changes in the extent of the Quelccaya Ice Cap, Peru. *Quat. Res.* **72,** 157–163 (2009).
10. Abbott, M. B. *et al.* Holocene paleohydrology and glacial history of the central Andes using multiproxy lake sediment studies. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **194,** 123–138 (2003).
11. Rodbell, D. T., Seltzer, G. O., Mark, B. G., Smith, J. A. & Abbott, M. B. Clastic sediment flux to tropical Andean lakes: records of glaciation and soil erosion. *Quat. Sci. Rev.* **27,** 1612–1626 (2008).
12. Jomelli, V. *et al.* Fluctuations of Andean tropical glaciers since the last millennium and palaeoclimatic implication: a review. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **281,** 269–282 (2009).
13. Rabatel, A., Francou, B., Jomelli, V., Naveau, P. & Grancher, D. A chronology of the Little Ice Age in the tropical Andes of Bolivia (16°S) based on moraines by lichenometry: implication for climate reconstruction. *Quat. Res.* **70,** 198–212 (2008).
14. Blard, P. H., Lavé, J., Pik, R., Wagnon, P. & Bourles, D. Persistence of full glacial conditions in the central Pacific until 15,000 years ago. *Nature* **449,** 591–594 (2007).
15. Wagnon, P., Ribstein, P., Francou, B. & Sicart, J. E. Anomalous heat and mass budget of Glaciar Zongo, Bolivia, during the 1997–98 El Niño year. *J. Glaciol.* **47,** 21–28 (2001).
16. Favier, V., Wagnon, P. & Ribstein, P. Glaciers of the outer and inner tropics: a different behavior but a common response to climatic forcing. *Geophys. Res. Lett.* **31,** L16403, doi:10.1029/2004GL020654 (2004).
17. Sicart, J. E., Hock, R. & Six, D. Glacier melt, air temperature and energy balance in different climates: the Bolivian Tropics, the French Alps, and northern Sweden. *J. Geophys. Res.* **113,** D24113, doi:10.1029/2008JD010406 (2008).
18. Braconnot, P., Marzin, C., Grègoire, L., Mosquet, E. & Marti, O. Monsoon response to changes in Earth's orbital parameters: comparisons between simulations of the Eemian and of the Holocene. *Clim. Past* **4,** 281–294 (2008).
19. Cruz, F. W. *et al.* Orbitally driven east–west antiphasing of South American precipitation. *Nature Geosci.* **2,** 210–214 (2009).
20. Baker, P. A. *et al.* The history of South American tropical precipitation for the past 25,000 years. *Science* **291,** 640–643 (2001).
21. Condom, T. *et al.* Transient modelling of lacustrine regressions: two case studies from the Andean Altiplano. *Hydrol. Process.* **18,** 2395–2408 (2004).
22. Garreaud, R., Vuille, M. & Clement, A. The climate of the Altiplano: observed current conditions and mechanisms of past changes. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **194,** 5–22 (2003).
23. Haug, G. H., Hughen, K. A., Sigman, D. M., Peterson, L. C. & Röhl, U. Southward migration of the intertropical convergence zone through the Holocene. *Science* **293,** 1304–1308 (2001).
24. Kienast, M. *et al.* Eastern Pacific cooling and Atlantic overturning circulation during the last deglaciation. *Nature* **443,** 846–849 (2006).
25. Kageyama, M. *et al.* Glacial climate sensitivity to different states of the Atlantic meridional overturning circulation: results from the IPSL model. *Clim. Past* **5,** 551–570 (2009).
26. Kull, C., Imhof, S., Grosjean, M., Zech, R. & Veit, H. Late Pleistocene glaciation in the Central Andes: temperature versus humidity control: a case study from the eastern Bolivian Andes (17°S) and regional synthesis. *Global Planet. Change* **60,** 148–164 (2008).
27. Francou, B., Vuille, M., Wagnon, P., Mendoza, J. & Sicart, J. E. Tropical climate change recorded by a glacier in the central Andes during the last decades of the twentieth century: Chacaltaya, Bolivia, 16-S. *J. Geophys. Res.* **108,** 4154, doi:10.1029/2002JD002959 (2003).
28. Bradley, R. S., Vuille, M., Diaz, H. F. & Vergara, W. Threats to water supplies in the tropical Andes. *Science* **312,** 1775–1755 (2006).
29. Urrutia, R. & Vuille, M. Climate change projections for the tropical Andes using a regional climate model: temperature and precipitation simulations for the end of the 21st century. *J. Geophys. Res.* **114,** D02108, doi:10.1029/2008JD011021 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Financial support was provided by the French INSU programme 'LEVE', the French ANR Escarcel programme no. 06-Vuln-010, UR Great Ice at IRD, and the French ANR El Paso programme no. 10-Blan-608-01. We thank P. Ginot, E. Perroy, E. Ramirez and the Bolivian authorities for facilitating our field work in the Zongo valley. The ^{10}Be measurements were performed at the ASTER AMS national facility (CEREGE, Aix en Provence), which is supported by INSU/CNRS, the French Ministry of Research and Higher Education, IRD and CEA. We are grateful to L. Léanni for handling the necessary physical and chemical preparations before the ^{10}Be measurements, and to M. Arnold, G. Aumaître and K. Keddadouche for help during the ^{10}Be AMS measurements. We acknowledge the international modelling groups for providing their data for analysis. The PMIP2/MOTIF Data Archive is supported by CEA and CNRS, and was funded by the EU project MOTIF (EVK2-CT-2002-00153) and the Programme National d'Etude de la Dynamique du Climate (PNEDC). The analyses were performed using model versions available on 23 May 2009 (http://pmip2.lsce.ipsl.fr/). Computing time to run the IPSL model was provided by CEA. We are grateful for the assistance received from IHH (Instituto de Hidraulica e Hídrologia) and UMSA (Universidad Mayor de San Andrés, La Paz).

Author Contributions V.J. and D.B. conducted the field work on Telata; M.K. and P.B. performed the GCM modelling; P.-H.B. developed the PDD numerical modelling. J.-E.S., P.W. and V.F. provided ablation and climatic data from the Zongo glacier and helped in developing the glacier mass balance model and with the energy balance interpretation. D.L.B. and R.B. participated in analysing and interpreting the cosmogenic data. V.J., M.K., V.F., D.B., D.G., M.-P.L., P.W., P.-H.B., J.-E.S., R.B., D.L.B. and M.V. interpreted the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to V.J. (jomelli@cnrs-bellevue.fr).

doi:10.1038/nature09984

Discovery of novel intermediate forms redefines the fungal tree of life

Meredith D. M. Jones[1,2], Irene Forn[3], Catarina Gadelha[4], Martin J. Egan[1,5], David Bass[2], Ramon Massana[3] & Thomas A. Richards[1,2]

Fungi are the principal degraders of biomass in terrestrial ecosystems and establish important interactions with plants and animals[1–3]. However, our current understanding of fungal evolutionary diversity is incomplete[4] and is based upon species amenable to growth in culture[1]. These culturable fungi are typically yeast or filamentous forms, bound by a rigid cell wall rich in chitin. Evolution of this body plan was thought critical for the success of the Fungi, enabling them to adapt to heterogeneous habitats and live by osmotrophy: extracellular digestion followed by nutrient uptake[5]. Here we investigate the ecology and cell biology of a previously undescribed and highly diverse form of eukaryotic life that branches with the Fungi, using environmental DNA analyses combined with fluorescent detection via DNA probes. This clade is present in numerous ecosystems including soil, freshwater and aquatic sediments. Phylogenetic analyses using multiple ribosomal RNA genes place this clade with *Rozella*, the putative primary branch of the fungal kingdom[1]. Tyramide signal amplification coupled with group-specific fluorescence *in situ* hybridization reveals that the target cells are small eukaryotes of 3–5 μm in length, capable of forming a microtubule-based flagellum. Co-staining with cell wall markers demonstrates that representatives from the clade do not produce a chitin-rich cell wall during any of the life cycle stages observed and therefore do not conform to the standard fungal body plan[5]. We name this highly diverse clade the cryptomycota in anticipation of formal classification.

To investigate phylogenetic diversity among the deepest branches of the fungal tree of life, we aligned a broad selection of fungal small subunit ribosomal DNA (SSU rDNA) sequences using published phylogenies[1] as a guide, to which we added a comprehensive sampling of environmental DNA sequences available in GenBank[6–8] (Supplementary Table 1). Our trees are consistent with previous analyses[1] demonstrating that nucleariids, a group of opisthokont amoebae[9], branch as sisters to the fungi, indicating a transition from a phagotrophic to an osmotrophic form very early within the fungal radiation[1]. Our analyses also recovered a highly diverse clade of environmental sequences branching with the fungi and demonstrated that current models of fungal evolution and biodiversity, which are largely based on cultured microbes, have missed a huge fraction of the kingdom (perhaps even approaching half). The analysis demonstrated preliminary bootstrap support (66/33%) for this large clade of environmental sequences branching with *Rozella*[10], the putative primary branch in the fungal phylogeny[1]. We name this clade cryptomycota (Fig. 1a).

The cryptomycota clade contains sequences recovered from diverse habitats and geographical locations, including soils, marine and freshwater sediments, freshwater planktonic samples and oxygen-depleted environments but it seems to be absent from samples of the upper marine water column (Fig. 1a, b). To investigate the ecology and cell biology of these deep-branching fungi, we designed several DNA probes for the detection of subgroups in the cryptomycota clade. Using the ARB program, with comprehensive GenBank and SILVA database sampling, we identified ten probes of about 18 base pairs (bp) that are specific to different sequences in the cryptomycota clade; probes and their target sequences are listed in Supplementary Table 2. Two probes were used successfully as forward PCR primers in combination with a general eukaryotic SSU rDNA reverse primer, 1520r (ref. 8; see Supplementary Table 2 and Fig. 1c). We then used PCR to test for the presence of the cryptomycota sequences termed CM1 and CM2 in multiple samples from a local freshwater pond, three freshwater reservoirs (Dartmoor National Park) and four coastal marine surface water samples (Devon, UK). Of the primer sequences tested, CM1 and CM2 consistently amplified cryptomycota rDNA from the Washington Singer pond (Exeter University, Devon, UK, 50.7339 °N, 3.5375 °W). We constructed clone libraries from both sets of amplicons and sequenced 12 clones from each, recovering only sequences that were 99% similar to Washington Singer CM1 in the first library and to the Lily Stem CM2 sequence previously sampled from Priest Pot pond (Cumbria, UK, 54.372 °N, 2.990 °W) in the second. This process demonstrated that both probes, when used as forward PCR primers, are specific to the two target groups in the Washington Singer pond samples. We did not detect either subgroup in the marine waters tested; however, only 0.8% of the thousands of eukaryotic environmental sequences retrieved from oceanic surface waters are classified as belonging to the Fungi[11], indicating a low density of fungi cells in the upper marine water column.

We then aimed to increase gene sampling so that we could perform additional multi-gene phylogenetic analyses to test the branching position of the environmental sequences. We constructed environmental gene libraries to sample the wider SSU–ITS1–5.8S–ITS2–large subunit (LSU) gene array, using the group-specific forward PCR primers in combination with a general eukaryotic LSU reverse primer, 28Sr1 (ref. 12), see Fig. 1c and Supplementary Table 2). This allowed a second round of phylogenetic analysis with increased character sampling and resulted in improved topology support values throughout the phylogeny (Supplementary Fig. 2). Importantly, CM1 and CM2 sequences clustered as a sister clade to *Rozella* with 96% support using two distinct bootstrap methods (shown in red on Fig. 1a and in full on Supplementary Fig. 2). This confirms that the environmental sequences branch with the genus *Rozella*.

To visualize the cells of the novel clade, we used the ten cryptomycota probe sequences detailed in Supplementary Table 2 as HRP-conjugated probes for tyramide signal amplification fluorescence *in situ* hybridization (TSA-FISH). Of these probes, CM1.1 and CM1.2 were consistently successful when we applied a high-stringency method combined with multiple helper probes[13,14]. We applied the TSA-FISH methods to the same water samples in which we had confirmed the presence of cryptomycota sequence using PCR (the Washington Singer pond and the Trenchford and Tottiford reservoirs sampled from Dartmoor National Park, UK). For both groups, our FISH analyses consistently recovered an ovoid cell type of 3–5 μm (Fig. 1d) among a dense 4′, 6-diamidino-2-phenylindole (DAPI)-stained microbial community ($n = 98$, CM1.1; $n = 58$, CM1.2). In contrast, the other eight

[1]School of Biosciences, University of Exeter, Exeter EX4 4QD, UK. [2]Department of Zoology, Natural History Museum, Cromwell Road, London SW7 5BD, UK. [3]Department of Marine Biology and Oceanography, Institut de Ciències del Mar, CSIC, Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain. [4]Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK. [5]Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston Massachusetts 02115, USA.

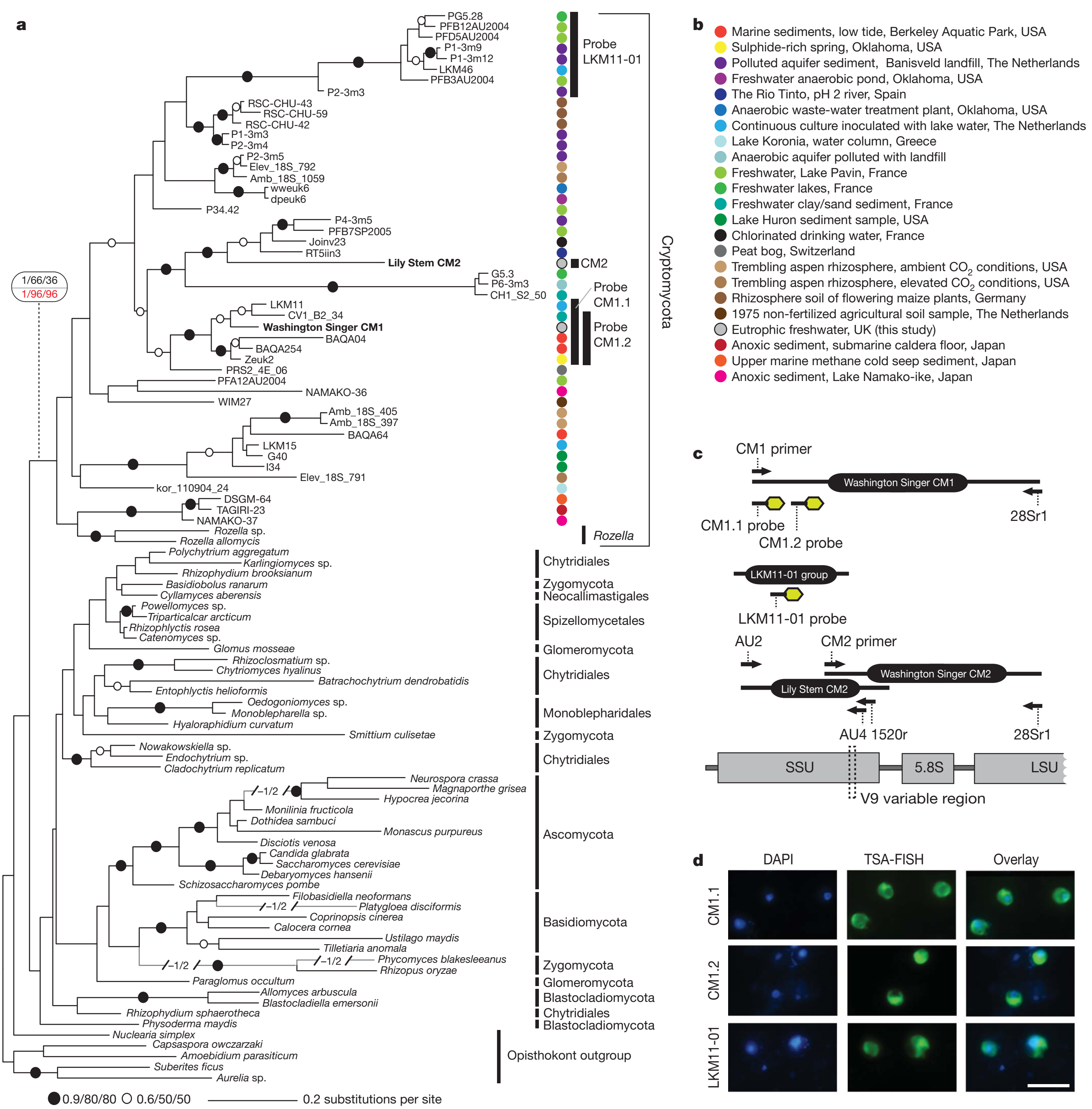


Figure 1 | Identification of the cryptomycota. **a**, Phylogeny demonstrates a diverse clade of environmental sequences (coloured dots) branching at the base of the Fungi. MrBayes tree topology was calculated from an alignment of 100 sequences and 1,012 DNA characters. Support values are summarized by black dots (indicating at least 0.9 Bayesian posterior probability and 80% bootstrap support by ML and Log-Det distance methods) or by black rings (Bayesian posterior probability above 0.6 and bootstrap support above 50%). The key node showing monophyly of cryptomycota encompassing *Rozella* is marked with actual values, with the SSU–5.8S–LSU support value (Supplementary Fig. 2a) shown in red. Shortened branches are in grey. Sequences targeted by the probes/primers are labelled on the tree. **b**, Provenance list of environmental DNA sequences. **c**, Schematic representation of the positions of primers (arrows) and probes (green hexagons) on a partial rDNA gene array (grey). **d**, TSA-FISH identification of cryptomycota cells (green) and DAPI staining of microbial community (blue). Scale bar, 10 µm.

cryptomycota group-specific probes and a panel of negative controls (detailed in the Methods) were consistently negative.

To complement this approach, we also used two LKM11 FISH probes previously used in association with additional eukaryotic probes to determine the eukaryotic community structure in freshwater environments[15]. LKM11 denotes a subsection of the environmental sequences branching within the cryptomycota clade detected in earlier studies[6–8]. We used the probe LKM11-01 (Fig. 1a, c) to identify a second subset of the cryptomycota clade and consistently recovered a cell type similar to that recovered for CM1 (Fig. 1d). The LKM11-02 probe was negative. We note that the abundance of cells identified by LKM11-01 was at least tenfold higher than observed for the CM1 cells, suggesting that the LKM11-01 subclade was more abundant than the CM1 subclade in the environments sampled.

To examine further the life cycle and morphology of the target groups in the freshwater samples, we tested whether the cells identified by

TSA-FISH possessed a flagellum. Flagella are a characteristic of chytrid fungi which, according to our phylogenetic analyses, branch closer to the cryptomycota than other fungal forms (Fig. 1a and Supplementary Figs 1 and 2). Co-localization of our TSA-FISH probes with the monoclonal antibody TAT1[16] against α-tubulin (a major structural component of eukaryotic flagella) was used to test for flagellate forms in the cryptomycota. Evidence of zoospore (flagellate spore) formation was observed for both the CM1 and LKM11-01 subgroups (Fig. 2a and Supplementary Fig. 3). In a comparative FISH experiment using 12 hybridizations from four samples (two from the Washington Singer eutrophic pond and one each from the oligotrophic Trenchford and Tottiford reservoirs, all with three replicates and using the LKM11-01 probe) we found that 47–85% of cells possessed a single flagellum ($n = 455$ of a total of 696 cells). These data demonstrate that the cells are present in a zoospore form, actively growing, reproducing and seeking food. During a typical chytrid cell cycle, a flagellum is usually lacking during the cyst phase when a cell wall is formed[5,17] and the cell enters dormancy; the majority of the remaining, unflagellate cryptomycota cells are therefore hypothesized to be cyst forms or cells in transition to form cysts (Fig. 2c).

We found preliminary evidence of a third phase of the cryptomycota life cycle. In both the LKM11-01 and CM1 groups, we observed non-flagellate cells attached to second-party cells (Fig. 2d, e) in what appeared to be an epibiotic (possibly parasitic or saprotrophic) association ($n = 39$ for LKM11-01, $n = 6$ for the CM1 probes). This represents a low total observation number ($n = 45$), an unavoidable consequence of sampling interacting microbes directly from the environment. However, in support of the interaction hypothesis, the cryptomycota cells were frequently found on diatoms. Furthermore, we saw examples of diatom cells with several attached cryptomycota cells (Supplementary Fig. 4): this would be unlikely if cell attachment were a sampling artefact. In addition, we observed non-illuminated cells attached to diatoms, indicating that the probes target cryptomycota specifically and that this was not an auto-fluorescence artefact affecting attached cells generally. Because cryptomycota are currently uncultured, we have to rely on environmental observations alone to understand their biology; it is therefore likely that several aspects of their life cycle have not yet been observed. Consequently, key cellular apparatuses may remain undetected, such as germ tubes, filter-feeding structures, rhizoids, hyphae, sporangia and cell division characteristics. However, our TSA-FISH observations of three life stages enable us to propose a skeleton life cycle for the cryptomycota lineages sampled, although we note that because of the evolutionary diversity of cryptomycota, this model is unlikely to represent accurately the numerous unobserved forms that branch within the wider clade (Fig. 2f).

The evolution of a chitin-rich cell wall was one of the most important acquisitions which drove the success and diversification of the Fungi[5], enabling these organisms to resist high osmotic pressure and feed by osmotrophy. *Rozella*, the only known genus branching within the cryptomycota clade (Fig. 1a), does not synthesize its own cell wall during many phases of its life cycle; instead, it appears to acquire cell

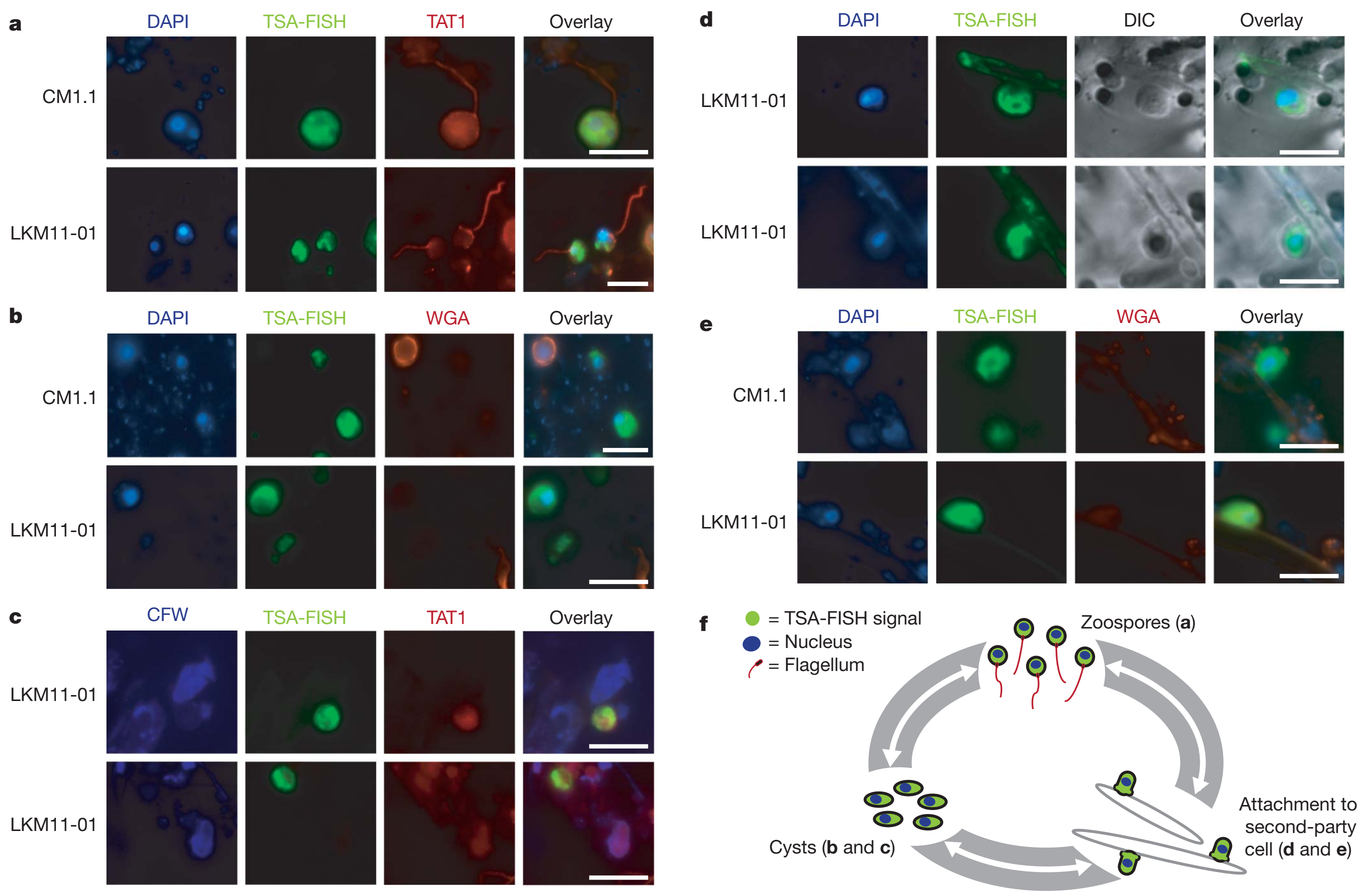


Figure 2 | Structural properties of cryptomycota cells and evidence for different life cycle stages. a, Micrographs showing flagella on cryptomycota cells, as detected by TAT1 tubulin antibody. **b**, Single cells without a chitin cell wall, as inferred by non-binding of wheat germ agglutinin (WGA). **c**, Non-flagellate putative cysts (lack of TAT1 signal indicates absence of flagellum), without a chitin/cellulose cell wall, as inferred by non-binding of calcofluor white. **d**, Cryptomycota cells attached to a second-party cell. Bright field differential interference contrast (DIC) shows filamentous structure of the second-party cell. **e**, Cryptomycota cells attached to a second-party cell; residual staining identifies the boundary of the second-party cell and WGA identifies absence of chitin during attachment. Scale bars, 10 µm. **f**, Putative cryptomycota skeleton life cycle (letters in brackets refer to micrographs **a**–**e**). This life cycle is limited to stages identified using TSA-FISH so additional stages are likely to remain unobserved (for example, sporangia stages and cell division) and the order of transition remains hypothetical. Furthermore, the diversity of the cryptomycota group strongly indicates that there are likely to be numerous life-cycle variations within the group, so this life cycle is unlikely to represent the wider diversity of cryptomycota.

wall material from its hosts during infection[18,19]. To investigate the cell biology of the cryptomycota further, it was therefore important to determine whether they produce a chitin- and cellulose-rich cell wall like other fungi (Supplementary Fig. 5), or whether this character is missing, consistent with the biology of *Rozella*[18,19].

We combined a TSA-FISH-based strategy with fluorescent stains specific to key compounds of the fungal cell wall. Calcofluor white binds and illuminates chitin and cellulose, the rigidifying components of fungal cell walls[5], and lectin wheat germ agglutinin binds and illuminates chitin[20]. Using these stains, we readily observed non-cryptomycota cells surrounded by chitin and chitin/cellulose in our samples (Fig. 2b, c) and in additional control experiments (Supplementary Fig. 5), confirming that these methodologies readily detect fungal cell walls. However, we did not observe a chitin/cellulose cell wall in any of the three cryptomycota life-cycle stages identified by FISH (Fig. 2b, c and e; Supplementary Figs 6 and 7). The TSA-FISH-based strategy and our samples may miss some specific stages of the life cycle with a chitin/cellulose-rich cell wall, yet our analyses show that the cryptomycota form non-flagellate putative cysts and attach to other cells (with the flagellum absent), all without a detectable chitin/cellulose cell wall. These results indicate that the key characteristic defining the kingdom Fungi — a growth and development strategy supported by a rigid chitin-rich cell wall[5] — is not present in key phases of the cryptomycota life cycle.

Using a combination of environmental DNA sequencing and fluorescence microscopy we have identified a new component of the fungal tree of life. We have tentatively named this wider group cryptomycota (crypto, hidden; mycota, of the kingdom Fungi) because a chitin-rich cell wall (one of the important fungal-defining characteristics) is so far undetected and because, with the exception of *Rozella*[18,19], the biology of this group is still largely cryptic. Significantly, the biodiversity within this clade is extensive, representing a breadth of rDNA molecular diversity similar to that of the currently sampled fungal kingdom.

METHODS SUMMARY

Phylogenetic analysis. The rDNA sequences from environmental samples and characterized species were obtained from GenBank, aligned using MUSCLE[21] and then manually refined and masked using SEAVIEW[22]. Phylogenies were calculated using 1,000 fast-ML[23] bootstraps, 1,000 Log-Det[24] distance bootstraps and Bayesian[25] methods.

FISH identification of cryptomycota cells. Surface water samples (3–4 litres) were collected from freshwater and marine coastal sites around Devon (Supplementary Table 3) and prefiltered with Miracloth (pore size 22–25 µm, Merck Chemicals). Half of the volume was immediately filtered under vacuum through a 2.0-µm polycarbonate membrane filter (Millipore) for DNA extraction. The remaining half was fixed with formaldehyde (3.7% final volume) overnight before filtration and storage at $-80\,^{\circ}$C. Oligonucleotide probes were designed against several cryptomycota subclades (Fig. 1c and Supplementary Table 2), conjugated with the HRP enzyme and used for TSA-FISH according to ref. 14. FISH optimization was performed using a range of stringency conditions (see online Methods).

To investigate whether a flagellum was present, filter pieces subjected to TSA-FISH hybridization were re-permeabilized with 0.1% v/v nonidet P-40 in PBS (10 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.2), blocked with 1% w/v bovine serum albumin in PBS and incubated for 1 h with the monoclonal antibody TAT1[16] against α-tubulin, followed by fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse immunoglobulins (Jackson ImmunoResearch/Stratech).

To determine the presence and composition of a cell wall, filter pieces prehybridized with TSA-FISH probes were counter-stained with cell wall markers: calcofluor white (Sigma, 1 mg ml^{-1}, fluorochrome for chitin and/or cellulose) and wheat germ agglutinin (Invitrogen, Alexa 555–580 tetramethylrhodamine conjugate, 5 µg ml^{-1}, for chitin). These were then viewed under epifluorescent ultraviolet radiation and red light respectively.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 July 2009; accepted 7 March 2011.
Published online 11 May 2011.

1. James, T. Y. *et al.* Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443,** 818–822 (2006).
2. Pirozynski, K. A. & Malloch, D. W. The origin of land plants: a matter of mycotrophism. *Biosystems* **6,** 153–164 (1975).
3. Wang, B. & Qiu, Y. L. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16,** 299–363 (2006).
4. Hawksworth, D. L. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycol. Res.* **105,** 1422–1432 (2001).
5. Bartnicki-Garcia, S. *Evolutionary Biology of the Fungi* (eds Rayner, A. D. M., Brasier, C. M. & Moore, D.) 389–403 (Cambridge University Press, 1987).
6. van Hannen, E. J., Mooij, W., van Agterveld, M. P., Gons, H. J. & Laanbroek, H. J. Detritus-dependent development of the microbial community in an experimental system: qualitative analysis by denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* **65,** 2478–2484 (1999).
7. Lepère, C., Domaizon, I. & Debroas, D. Unexpected importance of potential parasites in the composition of freshwater small-eukaryote community. *Appl. Environ. Microbiol.* **74,** 2940–2949 (2008).
8. Lefèvre, E. *et al.* Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: evidence from a molecular diversity study in a deep meromictic lake. *Environ. Microbiol.* **9,** 61–71 (2007).
9. Amaral Zettler, L. A., Nerad, T., O'Kelly, C. J. & Sogin, M. L. The nucleariid amoebae: more protists at the animal-fungal boundary. *J. Eukaryot. Microbiol.* **48,** 293–297 (2001).
10. Lara, E., Moreira, D. & Lopez-Garcia, P. Environmental clade LKM11 and *Rozella* form the deepest branching clade of Fungi. *Protist* **161,** 116–121 (2010).
11. Massana, R. & Pedrós-Alió, C. Unveiling new microbial eukaryotes in the surface ocean. *Curr. Opin. Microbiol.* **11,** 213–218 (2008).
12. Bass, D., Richards, T. A., Matthai, L., Marsh, V. & Cavalier-Smith, T. DNA evidence for global dispersal and probable endemicity of protozoa. *BMC Evol. Biol.* **7,** 162 (2007).
13. Fuchs, B. M., Glöckner, F. O., Wulf, J. & Amann, R. Unlabeled helper oligonucleotides increase the in situ accessibility to 16S rRNA of fluorescently labeled oligonucleotide probes. *Appl. Environ. Microbiol.* **66,** 3603–3607 (2000).
14. Kim, E. *et al.* Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proc. Natl Acad. Sci. USA* **108,** 1496–1500 (2011).
15. Mangot, J. F., Lepère, C., Bouvier, C., Debroas, D. & Domaizon, I. Community structure and dynamics of small eukaryotes targeted by new oligonucleotide probes: new insight into the lacustrine microbial food web. *Appl. Environ. Microbiol.* **75,** 6373–6381 (2009).
16. Woods, A. *et al.* Definition of individual components within the cytoskeleton of *Trypanosoma brucei* by a library of monoclonal antibodies. *J. Cell Sci.* **93,** 491–500 (1989).
17. Webster, J. & Weber, W. S. *Introduction to Fungi* 3rd edn (Cambridge University Press, 2007).
18. Held, A. A. The zoospore of *Rozella allomycis*: ultrastructure. *Can. J. Bot.* **53,** 2212–2232 (1975).
19. Held, A. A. *Rozella* and *Rozellopsis*: naked endoparasitic fungi which dress up as their hosts. *Bot. Rev.* **47,** 451–515 (1981).
20. Bulawa, C. E. Genetics and molecular biology of chitin synthesis in fungi. *Annu. Rev. Microbiol.* **47,** 505–534 (1993).
21. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5,** 113 (2004).
22. Galtier, N., Gouy, M. & Gautier, C. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **12,** 543–548 (1996).
23. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
24. Lockhart, P. J., Steel, M. A., Hendy, M. D. & Penny, D. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11,** 605–612 (1994).
25. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank: N. J. Talbot for advice, K. Gull for the TAT1 antibody, L. Guillou for access to curated SSU database and the Broad Institute of the Massachusetts Institute of Technology and Harvard for making their *Rhizopus* and *Batrachochytrium* genome sequence data publicly available. T.A.R. thanks the Leverhulme Trust for fellowship support. This work was primarily supported by a Natural Environment Research Council grant UK (NE/F011709/1). Additional support came from the Systematic Research Fund (awarded by the Systematics Association and the Linnean Society) to T.A.R., project FLAME (CGL2010-16304, MICINN) to R. M. and the BioMarKs project (European Funding Agencies from the ERA-net program BiodivERsA) to T.A.R. and R.M.

Author Contributions This study was conceived by T.A.R. and M.D.M.J. with assistance from D.B. and R.M.. M.D.M.J. performed the molecular biology experiments with assistance from I.F. (FISH), C.G. (immunolocalization) and M.J.E. (microscopy). T.A.R. performed the bioinformatics and phylogenetic analysis. T.A.R. and M.D.M.J. wrote the paper with assistance from D.B. and R.M.

Author Information Novel sequence data have been deposited in GenBank under accession numbers FJ687265, FJ687267 and FJ687268. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.A.R. (thomr@nhm.ac.uk).

METHODS

DNA extraction, PCR and environmental gene library construction. DNA was extracted from an epiphytic microbial community scraped from a submerged lily stem (Priest Pot pond, 54.372 °N, 2.990 °W) using the maximum yield protocol of the MoBio UltraClean Soil DNA Extraction Kit (Cambio), after filtration of a 50-ml stem-scraping suspended in Priest Pot water onto a GF-F filter (Whatman). Amplified PCR products of lily stem-scrapings (using the AU2 and AU4[26,27] primers, see Supplementary Table 2) were excised from the gel, purified (Wizard SV Gel and PCR Clean-Up System, Promega) and cloned (StrataClone, Stratagene) as suggested by the manufacturer. Plasmid purification (Wizard plus Miniprep Kit, Promega) was performed on 24 colonies before sequencing in the forward direction (Cogenics). Chromatograms were checked manually using Sequencher (Genecodes) and used for BLASTn searches of the National Center for Biotechnology Information, after which one positive cryptomycota clone was chosen for double-strand sequencing (GenBank number FJ687267).

Ten novel cryptomycota probes/primers (Supplementary Table 2), designed using the ARB[28] program (using database sampling from Silva[29] and GenBank[30]), were tested for specificity on community DNA extracted from the sites detailed in Supplementary Table 3 by using PCR with a forward-priming version of the probe in combination with the universal eukaryotic SSU reverse primer 1520r[8] (Supplementary Table 2). Two probes/primers were successful, yielding PCR products of the correct estimated amplicon length from Washington Singer environmental DNA using PCR amplification protocol 1 (Supplementary Table 4). The CM1–1520r PCR amplicon size was ~1,600 bp and the CM2–1520r PCR amplicon size was ~420 bp. Both amplicons crossed the variable V9 region (Fig. 1c). Two clone libraries were constructed and 12 clones from each library were sequenced.

Targeting longer rDNA sequences. We used the CM1 forward primer and an extended version of the CM2 forward primer in combination with the reverse LSU primer 28Sr1[12] (primer sequences given in Supplementary Table 2) to enable rRNA gene array PCR amplification. The CM1 and CM2 SSU–ITS1–5.8S–ITS2–LSU PCR amplifications were conducted using PCR protocols 2 and 3 respectively (Supplementary Table 4). Amplicons were purified, cloned as before and representatives of the CM1 (2,600 bp) and CM2 (1,449 bp) groups were double-strand sequenced. The Washington Singer CM2 group SSU-to-LSU sequence overlapped the Lily Stem CM2 SSU sequence by 316 identical bp, which encompassed the variable V9 region (Fig. 1c). For phylogenetic analyses, these sequences were concatenated, generating a sequence of 2,711 bp.

Sequence analysis, alignment and phylogenetic analyses. An SSU multiple-sequence DNA alignment was constructed using the work of James *et al.*[1,31] as a guide, with taxon sampling focusing on 'deep' branching fungal groups and on taxa with five rDNA sequence regions (SSU, ITS1, 5.8S, ITS2 and LSU) available (Supplementary Table 5). For the first alignment, we used the Washington Singer CM1 and Lily Stem CM2 SSU sequences as seeds for BLASTn searches of the GenBank non-redundant database to identify additional SSU sequences not sampled among currently published fungal phylogenies (for example, refs 1, 31). The alignment was calculated using the automatic alignment program MUSCLE[21], then manually refined and masked using SEAVIEW[22]. To check that we had included all available SSU sequences, we searched the GenBank non-redundant database again using all the members of the cryptomycota clade and the *Rozella* sequences as BLASTn search seeds (searches conducted May 2010). Many of the environmental sequences recovered were too short to be included in our phylogenetic analyses (Supplementary Table 1) but they indicated that the diversity of the cryptomycota clade is very extensive and even more ecologically varied than shown in Fig. 1a, b. The final phylogenetic analysis was based on environmental SSU sequences that overlapped both the CM1 and CM2 probe-binding regions; that is, only environmental SSU rDNA sequences of more than ~1,350 bp (BLAST searches conducted in May 2010). We removed all sequences with evidence of chimaeras identified by comparison to the curated SSU rDNA database (gift of L. Guillou).

A second multiple-sequence alignment was generated on the basis of the SSU–ITS1–5.8S–ITS2–LSU DNA sampling often used in the fungal tree of life project[1,31] and including our two cryptomycota sequences. This was processed, as described above, to generate a 54-sequence and 1,877-character data matrix with the ITS1 and ITS2 regions masked out because they were too variable for use. The alignment was checked for patterns of sequence variation that could be the products of chimaeras, both by visual analysis of signature sequences across the alignment and by constructing trees from one-sixth divisions of the masked alignment.

Both the SSU and SSU–5.8S–LSU sequence alignments were run through the program MODELGENERATOR[32] to identify the most appropriate model parameters for phylogenetic analysis (SSU: model of substitution GTR[33] + Γ (0.28, eight rate categories); SSU–5.8S–LSU: model of substitution GTR[33] + Γ (0.35, eight rate categories + I = 0.22)). Bayesian tree topologies were inferred for both DNA alignments using a Metropolis-coupled Markov chain Monte Carlo (MCMCMC) method. For this, we used MrBayes[25] with two runs, each with four Markov chains for 5,000,000 generations with default 'temperature' settings, a sampling frequency of 250 generations, six substitution categories (nst = 6) and an eight-category gamma model with spatial autocorrelation between rates at adjacent sites (rates = adgamma) and a covarion-like model (covarion = yes). The MCMCMC log likelihood results were compared and a 'burn-in' of 1,000 generations (SSU–5.8S–LSU analysis) and 2,500 generations (SSU analysis) was removed to sample only trees from the stationary phase of the MCMCMC searches. The resulting samples of trees were then used to construct the majority-rule consensus trees. Tree topologies were evaluated using two bootstrap methods. For each phylogeny, 1,000 bootstraps were calculated using the fast-ML method implemented in the PHYML program[23] using the model parameters from the MODELGENERATOR analyses. As compositional bias can cause phylogenetic reconstruction artefacts[34], we ran an additional bootstrap analysis using Log-Det methods[24] with stepwise addition (ten random starting trees per replicate) and a tree-bisection-reconnection branch-swapping algorithm. In both Log-Det bootstrap analyses, the 'proportion of invariant sites' parameter was estimated using PAUP[35]: SSU = 0.457 and SSU–5.8S–LSU = 0.509.

FISH identification of cryptomycota cells. Surface water samples (3–4 litres) were collected from freshwater and marine coastal sites around Devon (Supplementary Table 3) and prefiltered (pore size 22–25 µm) with Miracloth (Merck Chemicals). Half of the volume was immediately filtered under vacuum through a 2.0-µm polycarbonate membrane filter (Millipore) for DNA extraction. The remaining half was fixed with formaldehyde (3.7% final volume) overnight before filtration and storage at −80 °C. Oligonucleotide probes were designed against several cryptomycota subclades (Fig. 1a, c and Supplementary Table 2), conjugated with the HRP enzyme and used for TSA-FISH[13,14]. FISH optimization was performed using a range of stringency conditions (see below).

FISH stringency, helper probe, chitin wash and control experiments. The thirteen FISH probes, including a negative control probe (Supplementary Table 2), conjugated with the HRP enzyme, were used for TSA-FISH according to the method detailed in ref. 14. Initial hybridizations were at 35 °C with 30% formamide hybridization buffer. These were unsuccessful in revealing our target group. Formamide concentrations (30%, 40%, 45% and 50%) were varied in combination with a range of hybridization temperatures (35, 42 and 46 °C). Probes CM1.1 and CM1.2, both with a minimum of two mismatches to all non-target database sequences, were successful in illuminating candidate cryptomycota cells. These cells were not seen in the negative controls (detailed below).

To improve the rate of detection further, we designed additional helper probes (without the HRP conjugate) specific to the CM1.1 and CM1.2 target groups, which bind either side of the TSA-FISH probe-binding site (Supplementary Table 2). This approach should theoretically open up the tertiary structure of the ribosome and increase accessibility for the TSA-FISH probe to the target binding site[13,14]. Group-specific helper probes (Supplementary Table 2) were used in conjunction with the addition of 10% dextran sulphate to the hybridization reaction. For both probes, the greatest number of target cells was viewed using a 30% formamide hybridization buffer at 42 °C, with the addition of 10% dextran sulphate and helper probes.

We then combined our TSA-FISH protocol with a chitinase incubation step before hybridization, to test whether a chitin cell wall was inhibiting detection of the target cell[36]. Filter pieces were incubated in a 1 mg ml^{-1} solution of chitinase (Sigma) in modified PBS (10 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 5.5 rather than 7.2) with 1% SDS. Filters were incubated for 10 min at 30 °C, rinsed with distilled water and subjected to the TSA-FISH protocol as already described. The chitinase wash step did not appear to increase detection frequency and therefore a chitin cell wall was not considered an inhibitor of detection. The chitinase wash step was not used during further TSA-FISH analyses.

For the LKM11-01 probe, we used the methodology described in ref. 15 with filter pieces hybridized at a range of formamide concentrations (30%, 40%, 45% and 50%). The optimum concentration was 40% and this was used for all subsequent hybridizations. Again, for LKM11-01 TSA-FISH, six replicate preparations were completed using the chitinase wash step. These extra analyses again failed to increase detection frequency, indeed it often reduced the frequency of detection in replicate filters indicating that inhibition of TSA-FISH binding by a chitin cell wall was not a factor, and again this protocol step was dropped from further experiments.

Negative control tests were devised to check the specificity of the probes and eliminate the possibility of autofluorescence in the sample. Filter pieces from each sample were hybridized in identical conditions to those described above, but without a TSA-FISH probe, which was replaced with distilled water. Furthermore, hybridizations conducted using a TSA-FISH probe specific to a 16S plastid SSU rRNA gene (see Supplementary Table 2) for a novel algal group[14] and probes for further cryptomycota groups were used as additional negative controls. Positive cells were never observed.

Co-staining for flagella and cell wall apparatuses. To investigate whether a flagellum was present, filter pieces subjected to TSA-FISH hybridization were re-permeabilized with 0.1% v/v nonidet P-40 in PBS (10 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.2), blocked with 1% w/v bovine serum albumin in PBS and incubated for 1 h with the monoclonal antibody TAT1[16] against α-tubulin, followed by fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse immunoglobulins (Jackson ImmunoResearch/Stratech).

To determine the presence and composition of a cell wall, filter pieces pre-hybridized with TSA-FISH probes were counter-stained with cell wall markers: calcofluor white (Sigma, 1 mg ml^{-1}, fluorochrome for chitin and/or cellulose) and wheat germ agglutinin (Invitrogen, Alexa 555–580 tetramethylrhodamine conjugate, 5 μg ml^{-1}, for chitin). These were then viewed under epifluorescent ultraviolet radiation and red light respectively. These markers readily detected cell wall structures in a *Blastocladiella emersonii* culture (Supplementary Fig. 5) and among the total microbial population sampled on our environmental filters (Fig. 2b, c and Supplementary Figs 6 and 7).

26. Bass, D. *et al.* Yeast forms dominate fungal diversity in the deep oceans. *Proc. R. Soc. Lond. B* **274,** 3069–3077 (2007).
27. Vandenkoornhuyse, P., Baldauf, S. L., Leyval, C., Straczek, J. & Young, J. P. W. Extensive fungal diversity in plant roots. *Science* **295,** 2051 (2002).
28. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32,** 1363–1371 (2004).
29. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–7196 (2007).
30. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Res.* **34,** D16–20 (2006).
31. James, T. Y. *et al.* A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* **98,** 860–871 (2006).
32. Keane, T. M. *et al.* Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* **6,** 29 (2004).
33. Lanave, C., Preparata, G., Saccone, C. & Serio, G. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* **20,** 86–93 (1984).
34. Foster, P. G. & Hickey, D. A. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J. Mol. Evol.* **48,** 284–290 (1999).
35. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods).* Version 4 (Sinauer Associates, 2002).
36. Baschien, C., Manz, W., Neu, T. R. & Szewzyk, U. Fluorescence in situ hybrization of freshwater fungi. *Int. Rev. Hydrobiol.* **86,** 371–381 (2001).

LETTER

doi:10.1038/nature10131

Visual place learning in *Drosophila melanogaster*

Tyler A. Ofstad[1,2], Charles S. Zuker[1,2,3,4,5] & Michael B. Reiser[1]

The ability of insects to learn and navigate to specific locations in the environment has fascinated naturalists for decades. The impressive navigational abilities of ants, bees, wasps and other insects demonstrate that insects are capable of visual place learning[1–4], but little is known about the underlying neural circuits that mediate these behaviours. *Drosophila melanogaster* (common fruit fly) is a powerful model organism for dissecting the neural circuitry underlying complex behaviours, from sensory perception to learning and memory. *Drosophila* can identify and remember visual features such as size, colour and contour orientation[5,6]. However, the extent to which they use vision to recall specific locations remains unclear. Here we describe a visual place learning platform and demonstrate that *Drosophila* are capable of forming and retaining visual place memories to guide selective navigation. By targeted genetic silencing of small subsets of cells in the *Drosophila* brain, we show that neurons in the ellipsoid body, but not in the mushroom bodies, are necessary for visual place learning. Together, these studies reveal distinct neuroanatomical substrates for spatial versus non-spatial learning, and establish *Drosophila* as a powerful model for the study of spatial memories.

Vision provides the richest source of information about the external world, and most seeing organisms devote enormous neural resources to visual processing. In addition to visual reflexes, many animals use visual features to recall specific routes and locations, such as the placement of a nest or food source. When leaving the nest, bees perform structured 'orientation flights' to learn visual landmarks. If subsequently displaced from their outbound flight, bees take direct paths back to their nests using these learned visual cues[7]. However, it is not clear how insects, which have relatively compact nervous systems, perform these navigational feats. In mammals, the identification of place, grid and head direction cells suggests the existence of a 'cognitive map'[8]. Unfortunately, little is known about the cellular basis of invertebrate visual place learning. To identify the neurons and dissect the circuits that underlie navigation, we studied place learning in a genetically tractable model organism, *Drosophila melanogaster*.

To test explicitly for visual place learning in *Drosophila*, we developed a thermal–visual arena inspired by the Morris water maze[9] and a heat maze used with cockroaches and crickets[1,2] (Fig. 1a). In the *Drosophila* place learning assay, flies must find a hidden 'safe' target (that is, a cool tile) in an otherwise unappealing warm environment

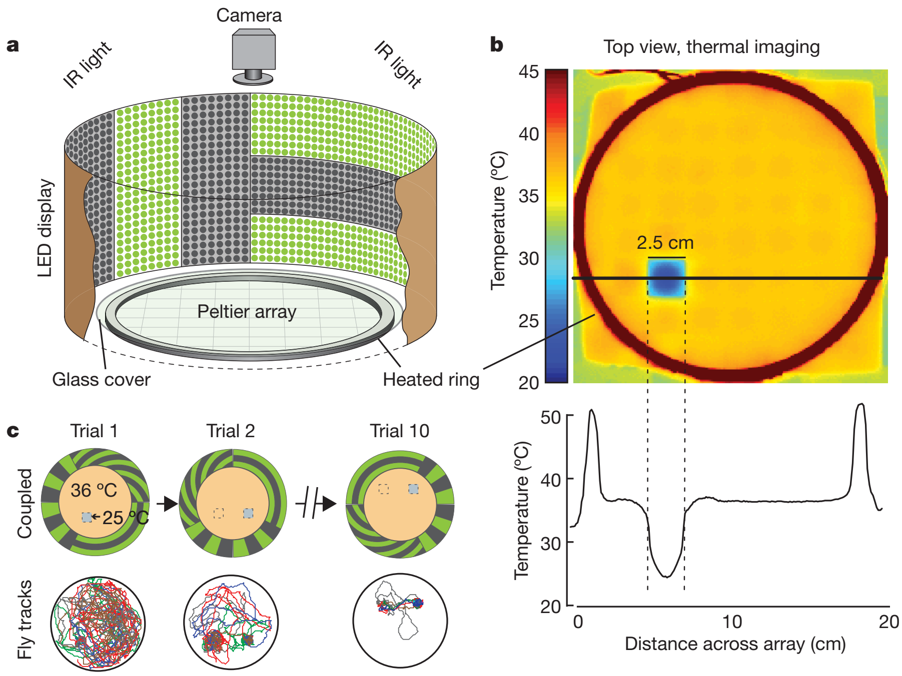


Figure 1 | *Drosophila* trained in a thermal–visual arena show place learning. **a**, Illustration of the arena. The floor is composed of 64 thermoelectric modules (a Peltier array), the panorama is provided by a 24 × 192 light-emitting-diode (LED) display and flies are recorded using a camera under infrared (IR) illumination. **b**, Top: thermal imaging view of the arena's floor showing the uniformly warm surface with a single cool tile; also shown is the heated ring barrier. Bottom: temperature readings across the arena. **c**, Trajectories of four representative flies from trials 1, 2 and 10 are shown below a diagrammatic representation of the visual panorama denoting the locations of the cool tile in the previous trial (dashed square) and in the current trial (blue square). In this coupled condition, the position of the cool tile relative to the visual panorama remains constant even as its absolute position changes between trials.

[1]Janelia Farm Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, Virginia 20147, USA. [2]Department of Neurosciences, Howard Hughes Medical Institute, University of California at San Diego, La Jolla, California 92093-0649, USA. [3]Department of Neurobiology, Howard Hughes Medical Institute, University of California at San Diego, La Jolla, California 92093-0649, USA. [4]Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA. [5]Department of Neuroscience, Howard Hughes Medical Institute, Columbia College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA.

(36 °C; Fig. 1b). Notably, there are no local cues that identify the cool tile. Rather, the only available spatial cues are provided by the surrounding electronic panorama that displays a pattern of evenly spaced bars in three orientations (Fig. 1a, c). To assay spatial navigation and visual place memory, fifteen adult flies are introduced in the arena and confined to the array surface by placing a glass disk on top of a 3-mm-high aluminium ring. During the first 5-min trial, nearly all flies (94%) eventually succeed in locating the cool target (data not shown). In subsequent trials, the cool tile and the corresponding visual panorama are rapidly shifted to a new location (rotated by either 90° clockwise or 90° anticlockwise, chosen at random). Importantly, the target and visual panorama are coupled so that although the absolute position of the cool tile changes, its location relative to the visual panorama remains constant (Fig. 1c). Our results (Fig. 2a, red trace, and Supplementary Movie 1) demonstrate that over the course of ten training trials flies improve dramatically in the time they require to locate the cool tile. This improvement is accomplished by taking a shorter (Fig. 2b), more direct route to the target (Fig. 2c), without noteworthy changes in the mean walking speed (Fig. 2d). To ensure that social interactions between flies were not influencing place learning (for example flies following each other to the safe spot), we also trained single flies and found that flies tested individually show equivalent place learning (Supplementary Fig. 1). As would be predicted for bona fide visual place learning, the improvement in place memory is critically dependent on the visual panorama. Flies tested in the dark show no improvement in the time, path length or directness of their routes to the target (Fig. 2, black traces).

To verify that flies are using the spatially distinct features of the visual panorama to direct navigation, we also tested flies using an uncoupled condition whereby the cool tile was still randomly relocated for each trial but the display remained stationary throughout. With this training regime, the visual panorama provides no consistent location cues, but idiothetic and weaker spatial cues such as the distance and local orientation of the arena wall are still available to the flies. Our results (Fig. 2, grey traces) demonstrate that flies trained with the uncoupled visual panorama show little improvement in the time taken to find the cool tile and no improvement in the directness of their approaches. Thus, spatially relevant visual cues are required for flies to learn the location of the target.

As a further test of visual place memory, flies were challenged immediately after training with a probe trial (trial 11) in which the visual landscape is relocated as usual but no cool tile is provided (to determine whether the flies will go to the non-existent safe spot). We proposed that if the flies learned to locate the cool tile by using the peripheral visual landmarks, then they should bias their searches to the area of the arena where the visual landscape indicates the cool tile should be, even when the target is absent. Indeed, flies preferentially search in the arena quadrant where they have been trained to locate the now 'imaginary' cool tile (Fig. 3, Supplementary Figs 2 and 3a, and Supplementary Movie 2). In contrast, if flies were trained in the dark or with an uncoupled visual landscape, conditions that contain no specific information about the location of the cool tile, the flies instead searched the arena uniformly during the probe trial (Fig. 3c, Supplementary Figs 2 and 3a, and Supplementary Movie 3). Together, these results demonstrate that fruit flies can learn spatial locations on the basis of distal visual cues and use this memory to guide navigation. By varying the time between the end of a single round of training (ten trials) and testing during a probe trial, we could also show that flies retain these visual place memories for at least 2 h (Fig. 3d).

We next considered where spatial memories are processed (or stored) in the *Drosophila* brain. We reasoned that specific regions of the fly brain would function as the neuroanatomical substrate for

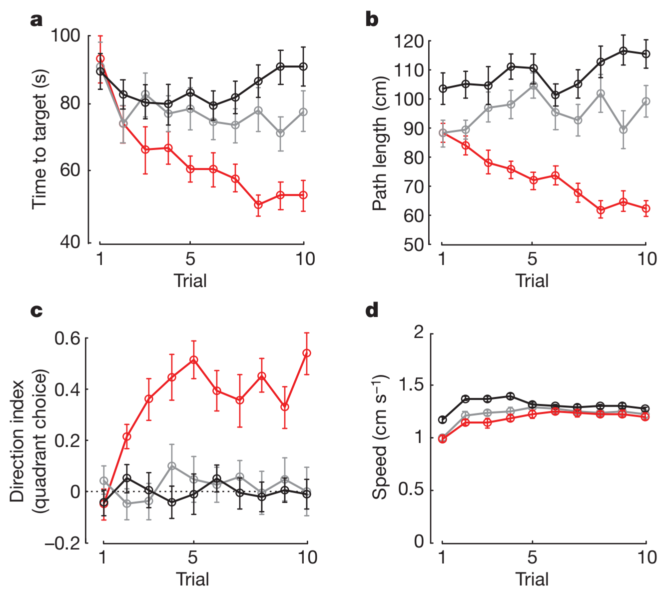


Figure 2 | Flies use visual cues to improve in place learning tasks. Flies were trained with a coupled visual panorama (red; $n = 33$ experiments, 495 flies), an uncoupled visual panorama (grey; $n = 21$ experiments, 315 flies) or in the dark (black; $n = 23$ experiments, 345 flies). **a**, When trained in the coupled condition (red) flies reduce the time taken to find the cool tile by nearly one-half, whereas flies trained with an uncoupled panorama (grey) or in the dark (black) show little or no reduction in the time taken to locate the target. **b**–**d**, The improvement seen for the coupled visual panorama is due to flies taking shorter (**b**), more direct (**c**) paths to the target rather than simply increasing walking speed (**d**). See Methods for details of calculations. Values shown are mean ± s.e.m.

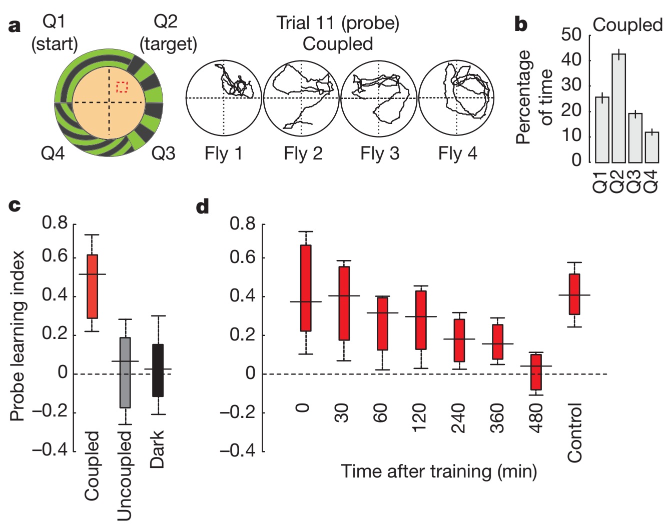


Figure 3 | Trained flies show a persistent search bias in the absence of the cool tile and retain this memory for several hours. Flies are tested in a probe trial (trail 11) in which the visual display is relocated but no cool tile is present. **a**, Trajectories from four representative flies, each plotted for 60 s after leaving its starting quadrant. Flies start in Q1; the dashed square denotes the 'expected' location of the cool tile, in Q2. **b**, Flies preferentially search in the quadrant where they have been trained to find the cool spot (Q2), even when the cool spot is absent. Values shown are mean ± s.e.m.; $n = 33$ experiments, 495 flies. **c**, Probe learning index is significantly greater than zero (indicating learning) when flies are trained with a coupled visual panorama (red; $P < 0.0001$, $n = 33$), but not when they are trained with an uncoupled visual panorama (grey; $P = 0.28$, $n = 21$) or in the dark (black; $P = 0.39$, $n = 23$). **d**, To test place memory retention, flies were tested in a probe trial at several time intervals following training ($n \geq 5$). Flies retain visual place memories for at least 2 h after training. Because flies were left in the arena between training and testing (for up to 8 h), 'control' refers to siblings placed in the arena for an equivalent amount of time (8 h) before training followed by immediate testing. Box plots indicate the median value (solid black line), 25th and 75th percentiles (box), and the data range (whiskers). For details of calculations and additional statistics, see Methods.

visual place learning, and we therefore set out to engineer and test animals in which different brain areas were selectively inactivated using the GAL4/UAS expression system. In essence, we conditionally silenced small subsets of neurons in adult flies by targeting expression of the inward-rectifying potassium channel Kir2.1[10]; to limit potential side-effects of Kir2.1 expression during development, we used a temperature-sensitive $GAL80^{ts}$ that blocks Kir2.1 expression when flies are reared at 18 °C but allows expression when the temperature is raised to 30 °C before testing[11]. *GAL4* driver lines were selected for expression in two areas: the mushroom bodies (Fig. 4a–c) and the central complex (Fig. 4d–f). The mushroom bodies have been the subject of extensive studies of learning and memory in *Drosophila*[12], and have been shown to be essential for associative olfactory conditioning[13] but not for some other forms of learning, such as tactile, motor and non-visually guided place learning[14–16]. The central complex is thought to be a site of orientation behaviour, multisensory integration and other 'high-order' processes[17,18]. In some social insects, the mushroom bodies have been implicated in visual place learning[19,20], and in the cockroach bilateral surgical lesions to these structures abolish spatial learning[1]. However, we see no evidence for involvement of the mushroom bodies in our assay. In fact, silencing mushroom body intrinsic neurons using the *GAL4* drivers R9A11, R10B08, R67B04 (Fig. 4a–c, g and Supplementary Fig. 3b) and OK107, or even by chemically ablating the mushroom bodies (using hydroxyurea[13]; Supplementary Fig. 4), had no significant effect on the performance of flies in visual place learning. The differing requirement for the mushroom bodies between *Drosophila* and other species may be explained by the observations that mushroom body inputs in *Drosophila* are predominantly olfactory[1,21]. In sharp contrast, silencing subsets of neurons with projections to the central complex ellipsoid body (lines R15B07 and R28D01) dramatically impaired visual place learning (Fig. 4d–h and Supplementary Fig. 3b). Notably, silencing a different subset of ring neurons with line R38H02 leaves visual place learning intact (Fig. 4f–h). Thus, specific circuits within the ellipsoid body (but not the entire structure) are necessary for visual place learning.

To confirm that silencing the ellipsoid body neurons in lines R15B07 and R28D01 produces a specific impairment in visual place memory, we tested these flies in a series of behavioural paradigms and showed they display normal locomotor, optomotor, thermosensory and visual pattern discrimination behaviours (Fig. 4i–l and Supplementary Fig. 5). In addition, we reasoned that if these flies have a general defect in memory (or in processing thermally driven learned behaviours), then they should show impairment in multiple types of learning (or in using thermal signals to drive learning and memory). Thus, we developed a novel olfactory conditioning paradigm using temperature (rather than electric shock[22]) as the unconditioned stimulus (Fig. 4m and Supplementary Fig. 6). As expected, silencing the mushroom bodies leads to a total loss of odour learning (Fig. 4m). In contrast, silencing subsets of neurons in the ellipsoid body has no effect on olfactory learning yet ablates visual place learning. Taken together, these results demonstrate that subsets of cells in the ellipsoid body are specifically required for visual place learning and substantiate the presence of distinct neuroanatomical substrates for visually guided spatial (place) versus non-spatial (olfactory) learning in *Drosophila*.

Mammals probably use place, grid and head direction cells to solve and perform navigational tasks[8]. The tight correlation between place

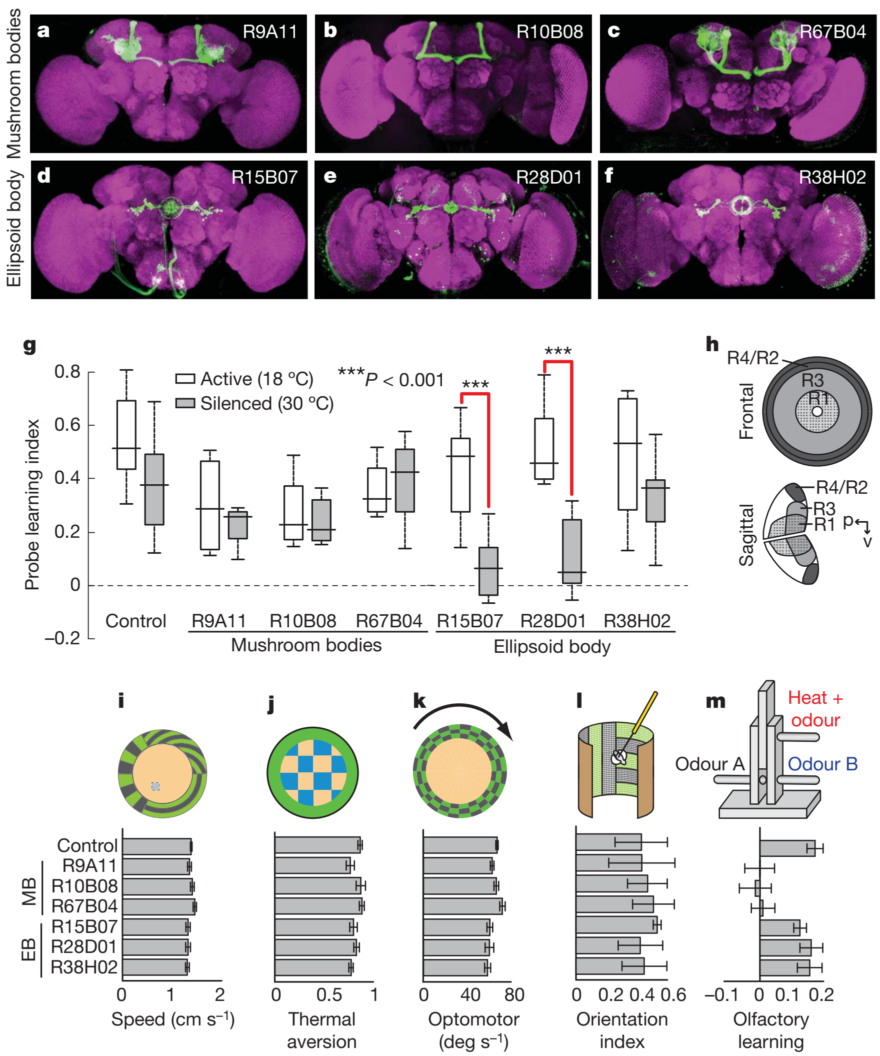


Figure 4 | Subsets of ellipsoid body ring neurons are required for place learning. **a**–**f**, *GAL4* driver lines targeting subsets of cells in the mushroom bodies (**a**–**c**) or the ellipsoid body (**d**–**f**); shown are the expression patterns for each driver using a green fluorescent protein reporter. **g**, White boxes indicate spatial learning before Kir2.1 induction; grey boxes indicate performance following Kir2.1 expression. Silencing ellipsoid body neurons projecting to R1 and R4 (R15B07) or to R1 alone (R28D01) severely impairs place learning ($P < 0.001$, red lines), whereas silencing mushroom body neurons or a separate subset of ellipsoid body neurons projecting to R4 alone (R38H02) leaves place learning intact. Box plots are as described in Fig. 3; $n \geq 8$ experiments. **h**, Schematic representation of ellipsoid body ring neuron anatomy. p, posterior; v, ventral. **i**–**m**, Flies with impaired place learning (expressing Kir2.1) show normal walking speed (**i**), heat aversion (**j**), optomotor response (**k**), visual pattern discrimination during tethered flight (**l**) and olfactory learning (**m**). In **i**–**k** ($n \geq 8$ experiments), values shown are mean ± s.e.m. In **l** ($n \geq 6$ flies) and **m** ($n = 8$ experiments), values shown are mean ± s.d. See Methods for details of calculations and additional statistical analysis. MB, mushroom body; EB, ellipsoid body.

cell activity and an animal's position in space has established the hippocampus as the substrate for a cognitive map. This map is probably informed by head direction cells (indicating an animal's orientation) and grid cells that tile the surrounding environment and could support path integration. Although it is not known whether there are direct correlates to these cells in flies, invertebrates are capable of solving similarly challenging navigational feats and do so using significantly smaller brains. Indeed, flies are able to use idiothetic cues, and path integration, to aid navigation[15,17,23,24]. Our studies now demonstrate that *Drosophila* can learn and recall spatial locations in a complex visual arena and do so with remarkable efficacy.

We also show that subsets of neurons in the fly brain (ring neurons of the ellipsoid body) are critical for visual place learning, probably by implementing, storing, or reading spatial information. Strikingly, flies in which we silenced ellipsoid body neurons have a basic 'circling' search routine (Supplementary Fig. 2d) that is reminiscent of the behaviour displayed by rats with hippocampal lesions[25]. Imaging of neuronal activity in the fly brain while the animal is executing a navigation task should help further define the role of the central complex, and ellipsoid body neurons in particular, in spatial memory (for example in a head-fixed preparation with a virtual-reality arena[26,27]). Ultimately, elucidating the cellular basis for place learning in *Drosophila* will help uncover fundamental principles in the organization and implementation of spatial memories in general.

Note added in proof: While our paper was under review, another study reported the use of a heat maze to study spatial search strategies in *Drosophila*[28].

METHODS SUMMARY

To control the thermal landscape, we developed an array of 64 1-inch-square, individually addressable thermoelectric modules arranged in an 8×8 grid (Fig. 1b). This array forms the floor of our test arena and is covered with black masking tape. To confine flies to this surface, a 3-mm-high, 8-inch-diameter, heated aluminium ring was placed around the outer perimeter of the arena and covered with a glass disk with a slippery surface. Visual cues were provided by a light-emitting-diode display positioned around the outer perimeter of the arena[29] (Fig. 1a). The experimental protocol included ten training trials (5 min each) followed by a probe trial (trial 11) in which the visual display was relocated in the absence of a cool spot. The navigational behaviour of the flies during a session was tracked off-line using CTRAX[30]. At the end of the experiment, flies were tested in the same arena for thermal preference and optomotor behaviour to confirm normal thermal and visual responses.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 October 2010; accepted 14 April 2011.

1. Mizunami, M., Weibrecht, J. M. & Strausfeld, N. J. Mushroom bodies of the cockroach: their participation in place memory. *J. Comp. Neurol.* **402,** 520–537 (1998).
2. Wessnitzer, J., Mangan, M. & Webb, B. Place memory in crickets. *Proc. R. Soc. B* **275,** 915–921 (2008).
3. Wehner, R. & Raber, F. Visual spatial memory in desert ants, *Cataglyphis bicolor* (Hymenoptera: Formicidae). *Experientia* **35,** 1569–1571 (1979).
4. Cartwright, B. A. & Collett, T. S. How honey bees use landmarks to guide their return to a food source. *Nature* **295,** 560–564 (1982).
5. Ernst, R. & Heisenberg, M. The memory template in *Drosophila* pattern vision at the flight simulator. *Vision Res.* **39,** 3920–3933 (1999).
6. Tang, S. & Guo, A. Choice behavior of *Drosophila* facing contradictory visual cues. *Science* **294,** 1543–1547 (2001).
7. Capaldi, E. A. & Dyer, F. C. The role of orientation flights on homing performance in honeybees. *J. Exp. Biol.* **202,** 1655–1666 (1999).
8. Moser, E. I., Kropff, E. & Moser, M. B. Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.* **31,** 69–89 (2008).
9. Morris, R. G. M. Spatial localization does not require the presence of local cues. *Learn. Motiv.* **12,** 239–260 (1981).
10. Baines, R. A., Uhler, J. P., Thompson, A., Sweeney, S. T. & Bate, M. Altered electrical properties in *Drosophila* neurons developing without synaptic transmission. *J. Neurosci.* **21,** 1523–1531 (2001).
11. McGuire, S. E., Mao, Z. & Davis, R. L. Spatiotemporal gene expression targeting with the TARGET and gene-switch systems in *Drosophila. Sci. STKE* **2004,** pl6 (2004).
12. Waddell, S. & Quinn, W. G. What can we teach *Drosophila*? What can they teach us? *Trends Genet.* **17,** 719–726 (2001).
13. de Belle, J. S. & Heisenberg, M. Associative odor learning in *Drosophila* abolished by chemical ablation of mushroom bodies. *Science* **263,** 692–695 (1994).
14. Wolf, R. *et al. Drosophila* mushroom bodies are dispensable for visual, tactile, and motor learning. *Learn. Mem.* **5,** 166–178 (1998).
15. Putz, G. & Heisenberg, M. Memories in *Drosophila* heat-box learning. *Learn. Mem.* **9,** 349–359 (2002).
16. Heisenberg, M. Mushroom body memoir: from maps to models. *Nature Rev. Neurosci.* **4,** 266–275 (2003).
17. Neuser, K., Triphan, T., Mronz, M., Poeck, B. & Strauss, R. Analysis of a spatial orientation memory in *Drosophila. Nature* **453,** 1244–1247 (2008).
18. Strauss, R. The central complex and the genetic dissection of locomotor behaviour. *Curr. Opin. Neurobiol.* **12,** 633–638 (2002).
19. Bernstein, S. & Bernstein, R. A. Relationships between foraging efficiency and size of head and component brain and sensory structures in red wood ant. *Brain Res.* **16,** 85–104 (1969).
20. Fahrbach, S. E. & Robinson, G. E. Behavioral development in the honey bee: toward the study of learning under natural conditions. *Learn. Mem.* **2,** 199–224 (1995).
21. Stocker, R. F. The organization of the chemosensory system in *Drosophila melanogaster*: a review. *Cell Tissue Res.* **275,** 3–26 (1994).
22. Tully, T. & Quinn, W. G. Classical conditioning and retention in normal and mutant *Drosophila melanogaster. J. Comp. Physiol. A* **157,** 263–277 (1985).
23. Wustmann, G., Rein, K., Wolf, R. & Heisenberg, M. A new paradigm for operant conditioning of *Drosophila melanogaster. J. Comp. Physiol. A* **179,** 429–436 (1996).
24. Zars, T. Spatial orientation in *Drosophila. J. Neurogenet.* **23,** 104–110 (2009).
25. Morris, R. G., Schenk, F., Tweedie, F. & Jarrard, L. E. Ibotenate lesions of hippocampus and/or subiculum: dissociating components of allocentric spatial learning. *Eur. J. Neurosci.* **2,** 1016–1028 (1990).
26. Seelig, J. D. *et al.* Two-photon calcium imaging from head-fixed *Drosophila* during optomotor walking behavior. *Nature Methods* **7,** 535–540 (2010).
27. Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L. & Tank, D. W. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neurosci.* **13,** 1433–1440 (2010).
28. Foucaud, J., Burns, J. G. & Mery, F. Use of spatial information and search strategies in a water maze analog in Drosophila melanogaster. *PLoS ONE* **5,** e15231 (2010).
29. Reiser, M. B. & Dickinson, M. H. A modular display system for insect behavioral neuroscience. *J. Neurosci. Methods* **167,** 127–139 (2008).
30. Branson, K., Robie, A. A., Bender, J., Perona, P. & Dickinson, M. H. High-throughput ethomics in large groups of Drosophila. *Nature Methods* **6,** 451–457 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We particularly thank M. Gallio for help with thermosensation and the development of temperature behavioural tests. We also thank G. Rubin for providing *GAL4* lines before publication, A. Jenett for their anatomical annotation and M. Dickinson for discussions and advice. Brain images were provided by the Janelia Fly Light Project. T. Laverty and the Janelia Fly Core assisted with *Drosophila* maintenance. Additional support was provided by J. Osborne, C. Werner, D. Olbris and M. Bolstad. We also thank V. Jayaraman, members of the Reiser and Zuker labs, Janelia Farm colleagues and the Janelia Fly Olympiad Project. This project was supported through the HHMI Janelia Farm Research Campus visitor programmed (T.A.O. and C.S.Z., hosted by M.B.R.). C.S.Z. is a HHMI investigator and a Senior Fellow at Janelia Farm.

Author Contributions All authors designed the study and wrote the manuscript. T.A.O. carried out the experiments and data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.S.Z. (cz2195@columbia.edu) or M.B.R. (reiserm@janelia.hhmi.org).

METHODS

Thermal–visual place learning arena. To control the thermal landscape, we developed an array of 64 1-inch-square, individually addressable thermoelectric modules arranged in an 8×8 grid (Oven Industries) (Fig. 1b). This array forms the floor of our test arena and is covered with black masking tape to create a uniform, featureless surface that can be replaced between experiments. Importantly, no thermal gradients exist that could guide flies to the cool spot from a distance (Fig. 1b and Supplementary Fig. 7), as confirmed by thermal imaging (Optotherm) and thermocouple measurements. Additionally, the absence of place learning in the dark and in the uncoupled condition (evidenced by near-zero direction and probe learning indices; Figs 2c and 3c) confirms that there are no significant non-visual or idiothetic cues in the arena that guide flies to the cool spot.

To confine flies to this surface, a 3-mm-high, 8-inch-diameter aluminium ring was placed around the outer perimeter of the arena and covered with a glass disk coated with a slippery silicon film (Sigmacote). To keep flies from walking on the walls, the aluminium ring was heated to $>50\,^{\circ}$C using insulated resistance wire (Pelican Wire, 29 AWG Nichrome 60 w/Kapton). Peripheral visual cues were provided using an electronically controlled light-emitting-diode (LED) display positioned around the outer perimeter of the arena[29] (Fig. 1a). In visual place learning trials, the LED panels were set to display a visual landscape composed of evenly spaced vertical, horizontal and diagonal bars. When viewed from the arena's centre, the width of each bar covered 15° of the LED display. When viewed from a distance of 8.8 inches, the maximum distance possible in our arena, each visual element subtends $\sim 8^{\circ}$ of the fly's visual field and should be easily resolvable by the fly. The entire arena is illuminated with infrared light (Smart Vision Lights) and fly activity was recorded with a Basler 622f CMOS camera fitted with an infrared passing filter.

Visual place learning protocol and analysis. The experimental protocol included ten training trials (5 min each) followed by a probe trial (trial 11) in which the visual display was relocated in the absence of a cool spot. At the end of the experiment, flies were tested in a temperature preference trial[31,32] and an optomotor trial[33] to measure normal thermal and visual responses. Flies were tracked offline using CTRAX fly-tracking software[30]. Fly centroid data were imported into MATLAB (Mathworks) and processed using custom scripts. In Fig. 2, the time to target, path length and speed are calculated, per fly, for the time window from the start of the trial until each fly reaches the target tile. Direction index (quadrant choice) is calculated as the number of flies that first enter the quadrant containing the cool tile (number correct) minus the number of flies that first pass into the opposite quadrant (number incorrect), divided by the total number of flies. To test whether flies show a bias for certain quadrants or rotation directions during training in the visual place learning arena, we tested for the dependence of the time to target on the target quadrant and on the rotation direction (clockwise or anticlockwise). This was accomplished by calculating the difference between the time to target for each trial (from each experiment) and the mean time to target for each trial across all experiments. The differences from the mean for training to quadrants 1, 2, 3 and or 4 are 1.7 ± 2.3, -0.5 ± 2.7, 0.0 ± 2.2 and -1.2 ± 2.2 s, respectively. The differences from the mean for clockwise and anticlockwise rotations are 0.6 ± 1.5 and -0.6 ± 1.8 s, respectively. Error is reported as s.e.m. For both tests, there are no statistically significant comparisons using one-way analysis of variance (ANOVA) at $P < 0.05$.

In Fig. 3b, the percentage of time spent in each of the quadrants was tested for statistical significance using one-way ANOVA with a Bonferroni correction for multiple comparisons. Flies spend significantly more time searching in Q2 than in Q1, Q3 or Q4 ($P < 0.01$). The probe learning index in Figs 3 and 4 is calculated from probe trial trajectories as the amount of time during the first 60 s after leaving the starting quadrant that flies spent searching in Q2 (the quadrant where they have been trained to locate the cool tile) minus the amount of time spent searching in Q4 (a quadrant that is the same distance from the starting quadrant, but in the wrong direction), divided by the total time spent in both quadrants. The P values reported in the legend of Fig. 3c were calculated using a one-tailed t-test. The probe learning index scores in Fig. 3d were tested for statistical significance ($P < 0.05$) using one-way ANOVA with a Bonferroni correction for multiple comparisons and pairwise comparisons to the uncoupled condition in Fig. 3c. The probe learning scores are significantly greater than uncoupled for up to 120 min after training. We note that although the probe learning index is not significantly different from the uncoupled control at 4 and 6 h, it is significantly greater than zero (one-tailed t-test, $P < 0.05$). Because flies were left in the arena between training and testing (for up to 8 h), 'control' refers to siblings placed in the arena for an equivalent amount of time (that is, 8 h) before training followed by immediate testing. The P values reported in the legend of Fig. 4 were calculated using a one-tailed t-test comparing place learning scores before (white box) and after (grey box) Kir2.1 induction. The probe learning index scores reported in Fig. 3c and 4g were also tested for statistical significance using one-way ANOVA with a Bonferroni correction for multiple comparisons. In Fig. 3c, flies trained in the coupled condition show significantly higher probe learning scores ($P < 0.01$) when compared with flies trained in the uncoupled condition or in the dark. There is no significant difference (at the $P < 0.05$ level) between uncoupled and dark. In Fig. 4g, R15B07 and R28D01 flies shifted to 30 °C are significant at $P < 0.01$ when compared with control flies. No other comparisons with control flies are significant at $P < 0.05$. In all box and whisker plots, the whiskers cover the range of the data, excluding outliers. Outliers are defined as data points greater than the 75th percentile of all data points plus 1.5 times the interquartile range ($q_3 + 1.5(q_3 - q_1)$) or data points less than the 25th percentile of all data minus 1.5 times the interquartile range ($q_1 - 1.5(q_3 - q_1)$). The majority of data sets presented as box plots contain no outlying data.

Following the probe trial, flies were tested for thermal preference by setting alternating tiles on the array of thermoelectric modules to either 25 or 36 °C. Flies were allowed to distribute for 2 min before the cool and warm tiles were switched. The flies were then allowed another 2 min to redistribute, and the thermal aversion index was calculated as the amount of time flies spent at 25 °C minus the amount of time spent at 36 °C, divided by the total time. Finally, flies were tested for normal optomotor responses by rotating a chequerboard pattern on the visual panorama clockwise and then anticlockwise at $90^{\circ}\,s^{-1}$ for 45 s. Optomotor responses are reported as the mean rotational speed (in the direction of the stimulus) of the flies over the course of these trials. No significant differences are observed in thermal aversion or optomotor response at $P < 0.05$ using one-way ANOVA with a Bonferroni correction for multiple comparisons.

Non-visual place learning. Work by a number of groups[15,17,23,24,34–36] has shown that flies can use idiothetic cues and path integration to navigate (in several forms of non-visual place memory task). To address whether idiothetic (that is, non-visual) cues are sufficient to guide navigation in the thermal–visual arena, we tested flies using a set of modified training protocols that included (a) keeping the cool tile stationary, (b) rotating the cool tile in a constant direction and (c) randomly rotating the cool tile between trials, all in the dark. Next we tested flies with (d) a stationary visual panorama and a randomly relocated cool tile, (e) a randomly rotating visual panorama and a stationary cool tile, and (f) random, independent relocations of the visual panorama and cool tile (Supplementary Fig. 8). We see no evidence of place learning when flies are trained and tested with any of these modified protocols. We note that flies may use idiothetic information while navigating; however, this experience is not sufficient for the formation of a place memory. To disperse flies from the stationary cool tile between trials, the entire array was heated to 36 °C for 60 s before the start of each trial. When trained with the standard coupled visual panorama (Figs 1 and 2), this manipulation does not impair visual place learning (data not shown).

Olfactory conditioning. Olfactory conditioning experiments were based on experiments using an elevated T-maze as described in ref. 22. The conditioning protocol was modified to use temperature rather than electric shock as the unconditioned stimulus (Fig. 4m and Supplementary Fig. 6). During conditioning, the training tube was heated to 36 °C concurrent with delivery of the first odour by passing a 5-V, 0.43-A current through a custom-built mesh of insulated resistance wire (Pelican Wire, 29 AWG Nichrome 60 w/Kapton) inserted into the training tube. Odours were delivered by bubbling an air stream through a vial containing odorant diluted in paraffin oil. Odours used were 5% 4-methylcyclohexanol (MCH; flow rate, $128\,ml\,min^{-1}$) and 5% 3-octanol (OCT; flow rate, $60\,ml\,min^{-1}$). Flow rate through training and testing tubes was normalized to $800\,ml\,min^{-1}$ by combining the odorant stream with a humidified clean air stream. About 200 flies were tested in each experiment, one half conditioned to MCH and the other conditioned to OCT. Learning indices were calculated as the average learning index of the two groups. All mushroom body lines (R9A11, R10B08 and R67B04) are significantly impaired in olfactory learning when compared with control flies ($P < 0.05$ using one-way ANOVA with a Bonferroni correction for multiple comparisons). No ellipsoid body lines are significantly different from control.

Tethered flight experiments. Closed-loop tethered flight experiments were performed as previously described using a cylindrical LED display and an optical wing-beat analyser to measure fly responses[29]. To test whether flies were capable of discriminating the visual features of the panoramic pattern in the visual place learning arena, we examined the orientation preference of flying flies for a flight arena pattern that was composed of four quadrants that display 15°-wide bar gratings, in either a vertical (quadrants 1 and 3) or horizontal (quadrants 2 and 4) direction. Each fly was allowed to selectively orient in a behavioural closed loop with this pattern for five trials of 50 s each, as part of an experimental series consisting of other closed- and open-loop trials, for which no further data are shown. Flies showed a clear preference for the vertical bars, so we quantified the behaviour with an orientation index that was calculated as the amount of time flies

oriented towards vertical bars minus the amount of time they oriented towards horizontal bars, divided by the total time. No significant differences are observed in the orientation index at $P < 0.05$ using one-way ANOVA with a Bonferroni correction for multiple comparisons.

Experimental animals. All flies used were female and, unless otherwise noted, are DL wild type. This strain is a laboratory culture produced by interbreeding dozens of wild caught isofemale lines, was established in 1995 and is maintained by Michael Dickinson's laboratory[37]. Flies were reared on standard media at 25 °C on a 16-h/8-h light/dark cycle. Visual place learning experiments were performed with 4-d-old adult flies during hours 11–15 of the flies' subjective day (where hour 0 corresponds to the transition from dark to light) in a room kept at 25 °C and 40% relative humidity. For neural silencing experiments, *w^+;tubP-GAL80^{ts};UAS-Kir2.1* flies (backcrossed ten generations into DL wild-type genetic background to control for the effects of genetic background[38] and known behavioural deficits with flies homozygous for *w^{1118}* (ref. 36)) were crossed to *GAL4* driver lines and reared at 18 °C. Two-day-old adult females were temperature-shifted to 30 °C for 40 h and then returned to 18 °C for 2 h before testing. *GAL4* driver lines were constructed as described in ref. 39 and were provided by Gerry Rubin. Control flies are *w^{1118};attP2* (the same genetic background as the *GAL4* lines) crossed to *w^+;tubP-GAL80^{ts};UAS-Kir2.1*. To ensure that backcrossing into the DL genetic background does not create conditions for PM hybrid dysgenesis (that is, mobilization of p-elements), our effectors/reporters (all marked with *mini-white*) are kept over balancer chromosomes and we regularly monitor for the appearance of the reporter in the wrong chromosome (an indication of transposition). Over the three years that we have been crossing and monitoring these stocks, we see no evidence of transposition. In Supplementary Fig. 5, line 78y was included to highlight locomotor abnormalities in a line with documented motor impairment[40].

31. Sayeed, O. & Benzer, S. Behavioral genetics of thermosensation and hygrosensation in Drosophila. *Proc. Natl Acad. Sci. USA* **93,** 6079–6084 (1996).
32. Zars, T. Two thermosensors in Drosophila have different behavioral functions. *J. Comp. Physiol. A* **187,** 235–242 (2001).
33. Strauss, R., Schuster, S. & Gotz, K. G. Processing of artificial visual feedback in the walking fruit fly Drosophila melanogaster. *J. Exp. Biol.* **200,** 1281–1296 (1997).
34. Wustmann, G. & Heisenberg, M. Behavioral manipulation of retrieval in a spatial memory task for Drosophila melanogaster. *Learn. Mem.* **4,** 328–336 (1997).
35. Zars, T., Wolf, R., Davis, R. & Heisenberg, M. Tissue-specific expression of a type I adenylyl cyclase rescues the rutabaga mutant memory defect: in search of the engram. *Learn. Mem.* **7,** 18–31 (2000).
36. Diegelmann, S., Zars, M. & Zars, T. Genetic dissociation of acquisition and memory strength in the heat-box spatial learning paradigm in Drosophila. *Learn. Mem.* **13,** 72–83 (2006).
37. Dickinson, M. H. Haltere-mediated equilibrium reflexes of the fruit fly, Drosophila melanogaster. *Phil. Trans R. Soc. B* **354,** 903–916 (1999).
38. de Belle, J. S. & Heisenberg, M. Expression of Drosophila mushroom body mutations in alternative genetic backgrounds: a case study of the mushroom body miniature gene (mbm). *Proc. Natl Acad. Sci. USA* **93,** 9875–9880 (1996).
39. Pfeiffer, B. D. *et al.* Tools for neuroanatomy and neurogenetics in Drosophila. *Proc. Natl Acad. Sci. USA* **105,** 9715–9720 (2008).
40. Martin, J. R., Raabe, T. & Heisenberg, M. Central complex substructures are required for the maintenance of locomotor activity in Drosophila melanogaster. *J. Comp. Physiol. A* **185,** 277–288 (1999).

doi:10.1038/nature10168

A genetically humanized mouse model for hepatitis C virus infection

Marcus Dorner[1], Joshua A. Horwitz[1], Justin B. Robbins[2], Walter T. Barry[1], Qian Feng[1], Kathy Mu[1], Christopher T. Jones[1], John W. Schoggins[1], Maria Teresa Catanese[1], Dennis R. Burton[2,3,4], Mansun Law[2], Charles M. Rice[1] & Alexander Ploss[1]

Hepatitis C virus (HCV) remains a major medical problem. Antiviral treatment is only partially effective and a vaccine does not exist. Development of more effective therapies has been hampered by the lack of a suitable small animal model. Although xenotransplantation of immunodeficient mice with human hepatocytes has shown promise, these models are subject to important challenges. Building on the previous observation that CD81 and occludin comprise the minimal human factors required to render mouse cells permissive to HCV entry *in vitro*[4], we attempted murine humanization via a genetic approach. Here we show that expression of two human genes is sufficient to allow HCV infection of fully immunocompetent inbred mice. We establish a precedent for applying mouse genetics to dissect viral entry and validate the role of scavenger receptor type B class I for HCV uptake. We demonstrate that HCV can be blocked by passive immunization, as well as showing that a recombinant vaccinia virus vector induces humoral immunity and confers partial protection against heterologous challenge. This system recapitulates a portion of the HCV life cycle in an immunocompetent rodent for the first time, opening opportunities for studying viral pathogenesis and immunity and comprising an effective platform for testing HCV entry inhibitors *in vivo*.

Humans and chimpanzees are the only species permissive to HCV infection. The basis for this highly restricted tropism is not completely understood, but may result from viral dependence on host factors present in only a few cell types. Murine cells are resistant to HCV entry, show inefficient viral replication, and may be blocked at later life cycle steps. HCV enters hepatocytes through the combined action of at least four host molecules: CD81 (ref. 1), scavenger receptor type B class I (SCARB1)[2], claudin 1 (CLDN1)[3] and occludin (OCLN)[4]. We have previously shown that CD81 and OCLN comprise the minimal human factors required for HCV uptake by rodent cells[4]. This led to the hypothesis that expression of these human orthologues could render mice susceptible to HCV infection *in vivo*. We therefore constructed recombinant adenoviruses encoding human CD81, SCARB1, CLDN1 and/or OCLN. Intravenous delivery of these vectors resulted in 100 to 1,000-fold overexpression of the corresponding messenger RNA in the murine liver and strong expression of all four proteins with the expected subcellular distribution (Supplementary Fig. 1). We determined that 18–25% of murine hepatocytes expressed human CD81 and OCLN together, and approximately 5% of cells expressed all four heterologous genes (Supplementary Fig. 2b–d). These results encouraged us to investigate infection of these animals. Unfortunately, HCV replication in mouse cells is inefficient *in vitro* and *in vivo*[5–9]. Consistent with this, challenge of mice expressing all four human factors with a firefly luciferase (Fluc)-encoding HCV genome (Jc1FLAG2(p7Fluc2A)) did not yield bioluminescent signal above background (Supplementary Fig. 3a). Direct measurement of Jc1FLAG2(p7Fluc2A) genome levels by polymerase chain reaction with quantitative reverse transcription (qRT–PCR) demonstrated a slight increase in HCV RNA in the serum (at 3 h) and liver (at 3 and 24 h); at 72 h, however, the signal was reduced to background (Supplementary Fig. 3b–d). These data highlight the difficulty of detecting HCV infection in cell types that do not support robust replication. In mouse cells, this defect may result from incompatibility between the viral replication machinery and murine factors and/or from exacerbated murine innate antiviral responses. Furthermore, adenoviral gene delivery strongly induces interferon-stimulated mouse genes, including viperin (also known as *Rsad2*), *Ifi44*, *Mx1*, *Oas1*, *Cxcl10* and *Eif2ak2* (also known as *PKR*), creating an environment that mimics recombinant IFN-α treatment (Supplementary Fig. 4) and may antagonize HCV replication[10].

As an alternative approach, we constructed a bicistronic HCV genome expressing CRE recombinase (Bi-nlsCre-Jc1FLAG2, abbreviated HCV-CRE), which activates a loxP-flanked luciferase reporter in the genome of the Gt(ROSA)26Sor$^{tm1(Luc)Kaelin}$ (Rosa26-Fluc) mouse[11]. Hydrodynamic delivery of HCV-CRE RNA into Rosa26-Fluc mice led to reporter signal in the liver, indicating that CRE recombinase is active in the context of the HCV genome (Supplementary Fig. 5). Delivery of a polymerase-defective HCV-CRE RNA produced similar results, indicating that significant CRE production was derived from initial translation without the need for replication (Supplementary Fig. 5). To test whether mice could be infected by authentic HCV particles, we generated Rosa26-Fluc animals expressing human CD81 and OCLN, or all four human entry factors, and inoculated these mice with cell-culture-derived HCV-CRE. In mice expressing all four transgenes, luciferase signals increased longitudinally, peaked at approximately 72 h after infection, and decreased thereafter; mice lacking the transgenes did not show significant reporter activity (Fig. 1b and Supplementary Fig. 6). All mice expressing at least human CD81 and OCLN could be successfully infected. Loss of signal after 72 h was probably due to strong antivector immunity, as evidenced by the increased frequencies of natural killer (NK) cells (Supplementary Fig. 7a); depletion of NK cells before adenovirus injection prolonged luminescence activity (Supplementary Fig. 7c). Bioluminescent signals were dependent on the doses of both adenovirus and HCV-CRE (Supplementary Fig. 6). HCV core, NS3 or NS5A could not be detected (data not shown), probably owing to inefficient viral replication in murine hepatocytes. To estimate the number of HCV-infected liver cells, we therefore used an indicator mouse strain in which CRE leads to activation of a nuclear-localized green fluorescent protein/β-galactosidase (GNZ) reporter (B6.129-Gt(ROSA)26Sor^{tm1Joe}/J, abbreviated Rosa26-GNZ)[12]. We observed HCV-CRE infection in approximately 4–5% of murine hepatocytes expressing human CD81 and OCLN (Fig. 1c and Supplementary Fig. 2h) and up to 20% expressing all four factors (Fig. 1c). We did not observe HCV uptake in non-parenchymal liver cells. These frequencies are consistent with previous estimates from primary human hepatocyte cultures[13,14] and HCV-positive patient tissues[15], although our system represents a single cycle infection and does not fully reflect the natural equilibrium. Taken together, these data provide— to our knowledge for the first time—evidence that mice can be rendered susceptible to HCV infection by a defined set of human genes.

[1]Center for the Study of Hepatitis C, The Rockefeller University, New York, New York 10065, USA. [2]Department of Immunology & Microbial Science, The Scripps Research Institute, La Jolla, California 92037, USA. [3]IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, California 92037, USA. [4]Ragon Institute of MGH, MIT and Harvard, Boston, Massachusetts 02129, USA.

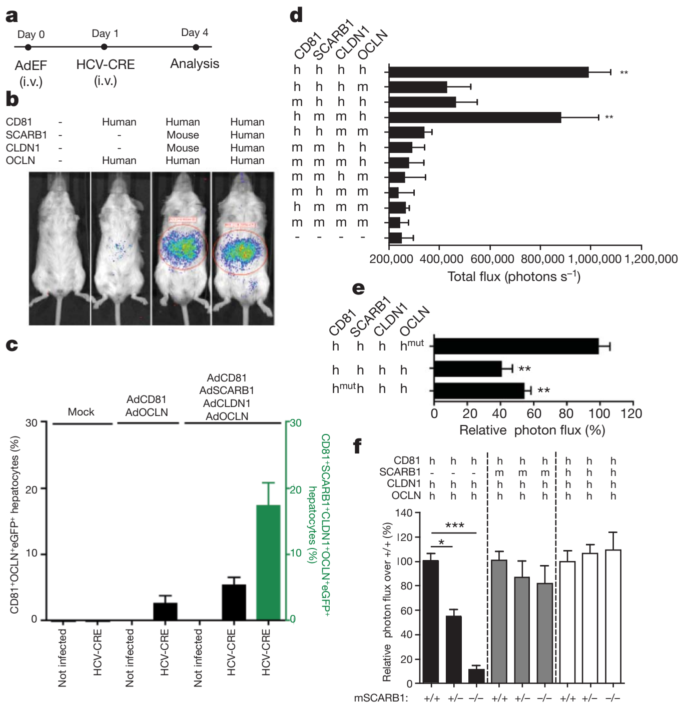


Figure 1 | Genetic requirements for HCV entry *in vivo*. **a**, Timeline for administration of adenovirus and HCV-CRE. **b**, HCV entry into Rosa26-Fluc mice 96 h after injection of medium ($n = 20$), 10^{11} adenovirus particles encoding CD81 and OCLN ($n = 10$) or 10^{11} particles encoding CD81, SCARB1, CLDN1 and OCLN of the indicated species ($n = 20$ each). Bioluminescence was measured at 72 h after injection of 2×10^7 50% tissue culture infectious doses ($TCID_{50}$) of HCV-CRE. **c**, Frequency of HCV-infected hepatocytes. Rosa26-GNZ mice were injected with medium or adenovirus encoding human CD81, mKate-SCARB1, Cerulean-CLDN1 and Venus/YFP-OCLN. Seventy-two hours after injection of 2×10^7 $TCID_{50}$ HCV-CRE, the frequency of infected hepatocytes was determined by flow cytometry. Data are per cent infected ($eGFP^+$) cells relative to either CD81/OCLN transduced (left axis) or CD81/SCARB1/CLDN1/OCLN transduced hepatocytes (right axis) **d**, Rosa26-Fluc mice were injected with human (h) and murine (m) entry factors 24 h before injection of HCV-CRE ($n \geq 4$). **e**, Entry factor mutants reduce infection *in vivo*. Combined delivery of human (h) and mutant human (h^{mut}), hCD81/F186L/E188K or hOCLN/mEL2 ($n = 3$). **f**, mSCARB1 is an essential HCV entry factor *in vivo*. Rosa26-Fluc mice were crossed with mSCARB1$^{-/-}$ mice and offspring was injected with adenoviruses encoding human CD81, SCARB1, CLDN1 and OCLN as indicated, and 24 h later with 2×10^7 $TCID_{50}$ HCV-CRE ($n = 5$–7). Data represent mean ± standard deviation (s.d.). Statistical significance was calculated by Kruskal–Wallis one-way analysis of variance (* $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$).

Previous studies in Chinese hamster ovary (CHO) and mouse fibroblast (NIH3T3) cells have shown that CD81 and OCLN are the primary determinants of HCV tropism at the level of entry[4]. To investigate the *in vivo* determinants of species restriction, we delivered combinations of human or mouse CD81, SCARB1, CLDN1 and OCLN to Rosa26-Fluc mice. In accordance with the *in vitro* data, HCV permissivity increased 6–10-fold in mice expressing human CD81 and OCLN, compared to those expressing only murine factors. Notably, addition of exogenous human or mouse SCARB1/CLDN1 (Fig. 1d) boosted permissivity compared to human CD81/OCLN alone (Fig. 1b), indicating that low endogenous entry factor levels may limit infection. Although overexpression of mouse CD81 or OCLN in combination with the other human entry factors permits low-level uptake, human CD81 and OCLN are required for optimal infection. To verify authentic co-receptor function, we delivered entry factors bearing mutations previously shown to disrupt their activities. CD81 mutations F186L and E188K are known to abrogate soluble E2 binding and to reduce HCV uptake by human hepatocytes[16]. Indeed, expression of mutant human CD81 reduced *in vivo* entry efficiency by about 50% (Fig. 1e). A human OCLN chimaera in which the second extracellular loop is replaced with the murine sequence has been shown to impair HCV entry *in vitro* by approximately fourfold[4]. Concordantly, HCV susceptibility of mice expressing the chimaeric molecule was reduced by approximately 50% (Fig. 1e). These data indicate that infection of murine hepatocytes *in vivo* is dependent on the native functions of human CD81 and OCLN.

Unlike the human-restricted entry factors, mouse SCARB1 seems to support HCV uptake *in vitro*[4] and *in vivo* (Fig. 1b). To validate the importance of this endogenous factor, we crossed mice with a disruption in *Scarb1* (SCARB1$^{-/-}$)[17] to the Rosa26-Fluc background. Expression of human CD81, CLDN1 and OCLN in mouse SCARB1$^{-/-}$ Rosa26-Fluc mice reduced HCV-CRE infection efficiency by ~90%, as compared to SCARB1$^{+/+}$, and ~50% compared to SCARB1$^{+/-}$ mice (Fig. 1f). This reduction seems to be a direct consequence of murine SCARB1 deficiency, and not transduction efficiency, as expression levels of CD81, CLDN1 and OCLN were similar in wild-type (Supplementary Fig. 1g) and SCARB1 knockout mice (Supplementary Fig. 8). Complementing mouse SCARB1$^{-/-}$ Rosa26-Fluc mice with human CD81, CLDN1 and OCLN and SCARB1 of either human or mouse origin fully restored infectivity (Fig. 1f). These data provide the first direct evidence that SCARB1 is a bona fide HCV entry factor *in vivo*. Taken together, these results indicate that infection of humanized mice occurs via authentic entry pathways, and highlight the value of this system for studying HCV co-receptor biology *in vivo*.

Despite considerable efforts, the lack of preventative or therapeutic HCV vaccines remains a major clinical challenge. Delivery of antibodies via passive immunization might be particularly suitable to block HCV entry, possibly limiting graft re-infection during liver transplantation and potentially boosting treatment success. Antibodies against E2 and CD81 have been shown to block HCV infection in cell culture[13,18] and human liver chimaeric mice[19,20]. We investigated passive immunization in the genetically humanized model. Delivery of anti-CD81 antibodies resulted in a dose-dependent inhibition of HCV-CRE infection, whereas isotype control immunoglobulins had no effect (Fig. 2a, d). Similarly, pre-incubation of HCV-CRE with antibodies directed against E2 (clone 3/11), but not an isotype control, significantly inhibited infection (Fig. 2b, e). These data further affirm that HCV is taken up in a glycoprotein-specific fashion *in vivo* and suggest the utility of this model for evaluation of passive immunization strategies.

Induction of broad-spectrum adaptive immunity is a key goal of HCV vaccine research. A promising candidate is a recombinant vaccinia virus (rVV) vector expressing HCV proteins, which has been shown to induce adaptive responses in rodents and chimpanzees[21]. We confirmed that HCV infection in our mouse model was observed across a panel of chimaeras expressing the structural proteins of diverse HCV genotypes (Fig. 3a). To evaluate this candidate in genetically humanized animals, Rosa26-Fluc mice were immunized intraperitoneally with 1×10^7 plaque forming units (p.f.u.) of rVV expressing the HCV proteins C-E1-E2-p7-NS2 (strain HCV-1, genotype 1a)[22]. Robust titres of anti-E2 antibodies (Fig. 3b) and decreased susceptibility to heterologous challenge with HCV-CRE expressing the structural proteins of genotypes 1b, 2a, or 4a (Fig. 3c) were observed. Protection directly correlated with the levels of anti-E2 antibodies in the serum. Challenge with genotype 1b (strain Con1) HCV-CRE virus, the envelope proteins of which are closely related to those of HCV-1, resulted in complete protection in two out of ten animals (Fig. 3d–f). Pooled sera of immunized—but not non-immune—mice neutralized diverse HCV genotypes *in vitro* (Supplementary Fig. 9) and reduced infection by 50% when transferred passively (Fig. 2c, f). These data demonstrate the value of the genetically humanized mouse as the first immunocompetent small animal model amenable to HCV challenge.

This study represents an important step forward in developing a robust small animal model for HCV infection and immunity. To our knowledge, this is the first time that any step in the viral life cycle has been recapitulated in a rodent simply by the expression of human genes. Previously developed liver chimaeric mice are susceptible to

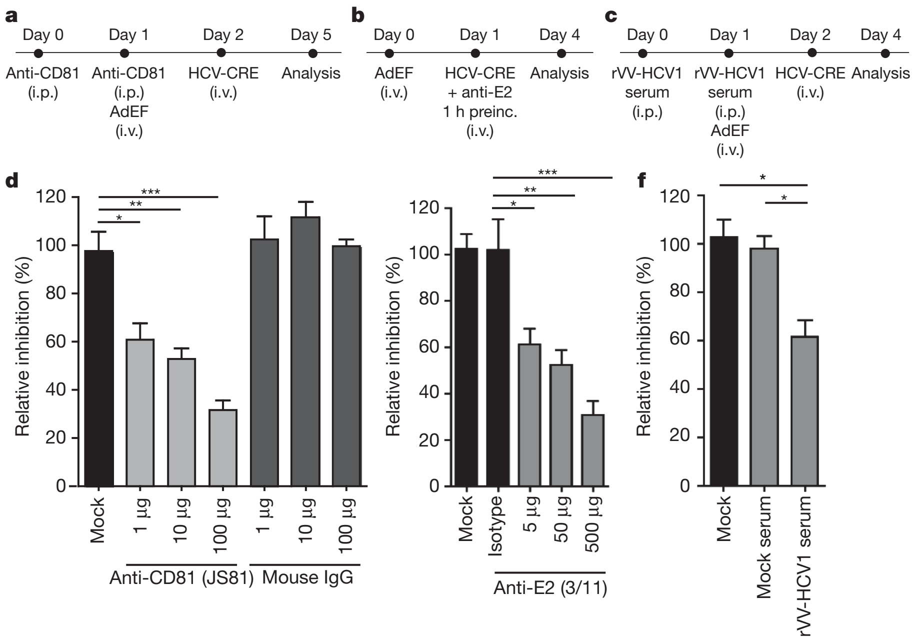


Figure 2 | HCV entry into murine hepatocytes *in vivo* can be blocked by antibodies or passive transfer of vaccine-induced antiserum. **a**–**c**, Experimental layout of CD81 blocking (**a**), virus pre-neutralization with anti-E2 (3/11) (**b**), and *in vivo* neutralization with pooled sera from immunized mice (**c**). i.p., intraperitoneal; preinc., preincubation. **d**, Rosa26-Fluc mice were injected with anti-CD81 antibodies (two doses of 1, 10 or 100 μg per animal at 24 h before and with injection of HCV-CRE; $n = 4$) or isotype control ($n = 3$). **e**, HCV-CRE was incubated with 5, 50 or 500 μg of anti-HCV E2 (clone 3/11) for 1 h before injection into Rosa26-Fluc mice expressing all four human entry factors ($n = 3$). **f**, Rosa26-Fluc mice were injected with 200 μl pooled serum from wild-type (FVB/NJ) mice immunized with rVV-HCV1 or naive control (two doses 24 h before and with injection of HCV-CRE; $n = 4$). Data shown are mean ± s.d. Statistical significance was calculated by Kruskal–Wallis one-way analysis of variance (* $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$).

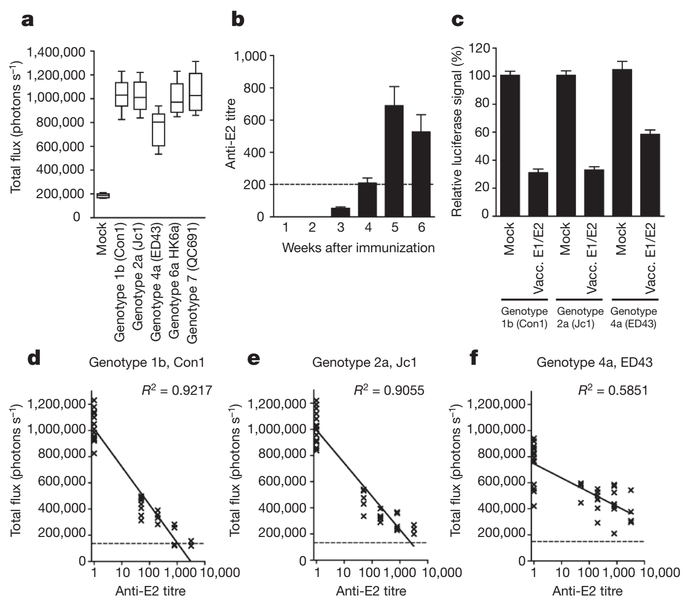


Figure 3 | Use of genetically humanized mouse model to evaluate vaccines against multiple HCV genotypes. **a**, Genetic humanization supports HCV entry mediated by structural proteins of various genotypes. Rosa26-Fluc mice expressing all four human entry factors were infected with intergenotypic chimaeras (Con1 genotype 1b, Jc1 genotype 2a, ED43 genotype 4a, HK6a genotype 6a and QC69 genotype 7a, 2×10^7 $TCID_{50}$ per animal; $n = 3$). **b**, Priming of humoral immune responses with rVV. Rosa26-Fluc mice were injected intraperitoneally with rVV encoding the HCV-1 (genotype 1a) structural genes (10^7 p.f.u. per animal; $n = 10$) and anti-HCV E2 antibody titres were determined by ELISA. **c**–**f**, Protection of rVV-immunized mice against challenge. Immunized mice were challenged with heterologous HCV strains Con1 (1b), Jc1 (2a) and ED43 (4a) and infection was quantified by bioluminescence imaging 72 h later ($n = 5$). Vacc., vaccinated. Data represent mean ± s.d.

HCV infection and support the entire viral life cycle[23–25]. Despite recent improvements[23], however, these models are hampered by low throughput, intra-experimental variability, and high costs[26]. Furthermore, the lack of a functional immune system negates their value for vaccination studies. The inbred mouse model presented here is not only susceptible to infection by diverse HCV genotypes, but it is also fully immunocompetent, providing the first small animal platform suitable for combined immunization and challenge studies. Although it may eventually be desirable to transgenically express the HCV entry factors, adenoviral expression allows rapid testing of various mutant proteins on diverse backgrounds. The genetically humanized mouse model provides proof of principle for such applications and holds promise for assessing the immunogenicity and efficacy of HCV vaccines.

METHODS SUMMARY

Mice. Gt(ROSA)26Sor$^{tm1(Luc)Kaelin}$ (ref. 11), B6;129-Gt(ROSA)26Sor^{tm1Joe}/J (ref. 27) and B6;129S2-Scarb1^{tm1Kri}/J (ref. 17) and FVB/NJ wild-type mice were obtained from The Jackson Laboratory. Mice were bred and maintained at the Comparative Bioscience Center of The Rockefeller University according to guidelines established by the Institutional Animal Committee.

Recombinant adenoviruses. Adenoviral constructs encoding human and murine homologues of the four HCV entry factors (CD81, SCARB1, CLDN1 and OCLN) and mutants thereof were created using the AdEasy Adenoviral Vector System (Agilent Technologies) according to the manufacturer's instructions.

Hepatitis C virus. Plasmids encoding chimaeric HCV genomes, including Jc1FLAG2(p7Fluc2A) and bicistronic HCV genomes expressing CRE, were linearized with XbaI and transcribed using MEGAscript T7 (Ambion). RNA was electroporated into Huh-7.5 cells using an ECM 830 electroporator (BTX Genetronics) and infectious virus was collected from supernatants 48–72 h after transfection.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 November; accepted 28 April 2011.

1. Pileri, P. *et al.* Binding of hepatitis C virus to CD81. *Science* **282,** 938–941 (1998).
2. Scarselli, E. *et al.* The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J.* **21,** 5017–5025 (2002).

3. Evans, M. J. *et al.* Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature* **446,** 801–805 (2007).
4. Ploss, A. *et al.* Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature* **457,** 882–886 (2009).
5. Zhu, Q., Guo, J. T. & Seeger, C. Replication of hepatitis C virus subgenomes in nonhepatic epithelial and mouse hepatoma cells. *J. Virol.* **77,** 9204–9210 (2003).
6. Uprichard, S. L., Chung, J., Chisari, F. V. & Wakita, T. Replication of a hepatitis C virus replicon clone in mouse cells. *Virol. J.* **3,** 89 (2006).
7. Chang, K. S. *et al.* Replication of hepatitis C virus (HCV) RNA in mouse embryonic fibroblasts: protein kinase R (PKR)-dependent and PKR-independent mechanisms for controlling HCV RNA replication and mediating interferon activities. *J. Virol.* **80,** 7364–7374 (2006).
8. Lin, L. T. *et al.* Replication of subgenomic hepatitis C virus replicons in mouse fibroblasts is facilitated by deletion of interferon regulatory factor 3 and expression of liver-specific microRNA 122. *J. Virol.* **84,** 9170–9180 (2010).
9. McCaffrey, A. P. *et al.* Determinants of hepatitis C translational initiation *in vitro*, in cultured cells and mice. *Mol. Ther.* **5,** 676–684 (2002).
10. Schoggins, J. W. *et al.* A diverse array of gene products are effectors of the type I interferon antiviral response. *Nature* **472,** 481–485 (2011).
11. Safran, M. *et al.* Mouse reporter strain for noninvasive bioluminescent imaging of cells that have undergone Cre-mediated recombination. *Mol. Imaging* **2,** 297–302 (2003).
12. Awatramani, R., Soriano, P., Mai, J. J. & Dymecki, S. An Flp indicator mouse expressing alkaline phosphatase from the *ROSA26* locus. *Nature Genet.* **29,** 257–259 (2001).
13. Ploss, A. *et al.* Persistent hepatitis C virus infection in microscale primary human hepatocyte cultures. *Proc. Natl Acad. Sci. USA* **107,** 3141–3145 (2010).
14. Jones, C. T. *et al.* Real-time imaging of hepatitis C virus infection using a fluorescent cell-based reporter system. *Nature Biotechnol.* **28,** 167–171 (2010).
15. Liang, Y. *et al.* Visualizing hepatitis C virus infections in human liver by two-photon microscopy. *Gastroenterology* **137,** 1448–1458 (2009).
16. Higginbottom, A. *et al.* Identification of amino acid residues in CD81 critical for interaction with hepatitis C virus envelope glycoprotein E2. *J. Virol.* **74,** 3642–3649 (2000).
17. Rigotti, A. *et al.* A targeted mutation in the murine gene encoding the high density lipoprotein (HDL) receptor scavenger receptor class B type I reveals its key role in HDL metabolism. *Proc. Natl Acad. Sci. USA* **94,** 12610–12615 (1997).
18. Catanese, M. T. *et al.* High-avidity monoclonal antibodies against the human scavenger class B type I receptor efficiently block hepatitis C virus infection in the presence of high-density lipoprotein. *J. Virol.* **81,** 8063–8071 (2007).
19. Meuleman, P. *et al.* Anti-CD81 antibodies can prevent a hepatitis C virus infection *in vivo*. *Hepatology* **48,** 1761–1768 (2008).
20. Law, M. *et al.* Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge. *Nature Med.* **14,** 25–27 (2008).
21. Youn, J. W. *et al.* Evidence for protection against chronic hepatitis C virus infection in chimpanzees by immunization with replicating recombinant vaccinia virus. *J. Virol.* **82,** 10896–10905 (2008).
22. Ralston, R. *et al.* Characterization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia viruses. *J. Virol.* **67,** 6753–6761 (1993).
23. Bissig, K. D. *et al.* Human liver chimeric mice provide a model for hepatitis B and C virus infection and treatment. *J. Clin. Invest.* **120,** 924–930 (2010).
24. Mercer, D. F. *et al.* Hepatitis C virus replication in mice with chimeric human livers. *Nature Med.* **7,** 927–933 (2001).
25. Meuleman, P. *et al.* Morphological and biochemical characterization of a human liver in a uPA-SCID mouse chimera. *Hepatology* **41,** 847–856 (2005).
26. de Jong, Y. P., Rice, C. M. & Ploss, A. New horizons for studying human hepatotropic infections. *J. Clin. Invest.* **120,** 650–653 (2010).
27. Stoller, J. Z. *et al.* Cre reporter mouse expressing a nuclear localized fusion of GFP and β-galactosidase reveals new derivatives of Pax3-expressing precursors. *Genesis* **46,** 200–204 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Sable, E. Castillo, A. Forrest, M. Panis, S. Pouzol, S. Shirley, A. Webson and E. Giang for laboratory support, L. Chiriboga and H. Yee for technical assistance, J. Bukh and Apath, LLC for providing the prototype intergenotypic HCV chimaeras and C. Murray for editing the manuscript. This study was supported in part by award number RC1DK087193 (to C.M.R. and A.P.) from the National Institute of Diabetes and Digestive and Kidney Diseases, R01AI072613 (to C.M.R.), R01AI079031 (to M.L.) and R01AI071084 (to D.R.B.) from the National Institute for Allergy and Infectious Disease, The Starr Foundation and the Greenberg Medical Institute. M.D. was supported by a postdoctoral fellowship from the German Research Foundation (Deutsche Forschungsgesellschaft) and M.T.C. by funds from The Rockefeller University's Women & Science Fellowship Program. J.W.S. and C.T.J are recipients of Ruth L. Kirschstein National Research Service Awards from the National Institute of Health (F32DK082155 to J.W.S., F32DK081193 to C.T.J.).

Author Contributions M.D., C.M.R. and A.P. designed the project, analysed results and wrote the manuscript. M.D., J.A.H., J.B.R., W.T.B., Q.F., K.M., M.T.C. and M.L. performed the experimental work, J.W.S., C.T.J. and D.R.B. provided reagents.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.P. (aploss@rockefeller.edu).

METHODS

Animals and cell lines. Gt(ROSA)26Sor$^{tm1(Luc)Kaelin}$ (ref. 11) (Rosa26-Fluc), B6;129-Gt(ROSA)26Sor^{tm1Joe}/J (ref. 27) (Rosa26-GNZ), B6;129S2-Scarb1^{tm1Kri}/J (ref. 17) (SCARB1$^{-/-}$) and FVB/NJ (wild-type) mice were obtained from The Jackson Laboratory. Rosa26-Fluc mice contain the firefly luciferase (luc) gene inserted into the Gt(ROSA)26Sor locus. Expression of the luciferase gene is blocked by a loxP-flanked STOP fragment placed between the luc sequence and the Gt(ROSA)26Sor promoter. CRE recombinase mediated excision of the transcriptional stop cassette results in luciferase expression in Cre-expressing tissues. Rosa26-GNZ knock-in mice have widespread expression of a nuclear-localized green fluorescent protein/β-galactosidase fusion protein (GFP-NLS-GNZ) once an upstream loxP-flanked STOP sequence is removed. When CRE recombinase is introduced into cells the resulting GNZ fusion protein expression allows for enhanced (single-cell level) visualization. Mice were bred and maintained at the Comparative Bioscience Center of the Rockefeller University according to guidelines established by the Institutional Animal Committee. Huh-7.5, Huh-7.5.1, 293T, HepG2 and HEK293 were maintained in DMEM with 10% fetal bovine serum (FBS) and 1% nonessential amino acids (NEAA) and 786-O cells were maintained in RPMI with 10% FBS and 1% NEAA.

Antibodies. Blocking antibodies against CD81 (JS81) and IgG1 control antibodies were obtained from BD Biosciences. Antibodies against NS5A[28] and E2 (clone 3/11)[29] have been described previously. Antibodies for the detection of human CD81 were purchased from BD Biosciences, OCLN from BD Biosciences (for histology) and from Invitrogen (for western blotting), CLDN1 from Invitrogen (for western blotting) and Abcam (for histology), and SCARB1 from Genetex (for histology) and from BD Biosciences (for western blotting). Antibodies for *in vivo* depletion of T cells (anti-mouse CD4 and anti-mouse CD8) were obtained from Biolegend and the antibody for depletion of NK cells (anti-mouse Asialo GM1) was purchased from Cedarlane. Fluorochrome-conjugated antibodies against mouse CD3, CD4, CD8, CD49b, Ly-6G were obtained from BD Biosciences and the IRDye700-conjugated anti-Luciferase antibody was acquired from Acris antibodies and Bethyl Laboratories.

Adenovirus constructs. Adenoviral constructs encoding human and murine homologues of the four HCV entry factors (CD81, SCARB1, CLDN1 and OCLN) and fluorescently tagged human SCARB1, CLDN1 and OCLN were created using the AdEasy Adenoviral Vector System (Agilent Technologies) according to the manufacturer's instructions. Briefly, human and murine entry factor cDNA was PCR-amplified from TRIP-based constructs[4] and inserted into the pShuttle-CMV using KpnI/XhoI sites (KpnI/NotI for human SCARB1). Mouse Kate-SCARB1, Cerulean-CLDN1, and Venus-OCLN fusions were cut from existing TRIP-based constructs and inserted into pShuttle-CMV using compatible restriction sites. Recombinant pShuttle-CMV plasmids were linearized with PmeI and ligated to pAdEasy by homologous recombination followed by electroporation into BJ5183 cells (Agilent). Recombinant pShuttle-pAdEasy constructs were identified by PacI restriction analysis. All plasmid constructs were verified by DNA sequencing.

Primer sequences were as follows. Human *CD81*, forward primer, CGCGGTACCCCACCATGGGAGTGGAGGGCTGCAC; reverse primer, ACGCTCGAGTCAGTACACGGAGCTGTTCC. Mouse *Cd81*, forward, GCATACAGGGTACCGCCACCATGGGAGTGGAGGGCTGCACCAAA; reverse, GCATACAGCTCGAGTCAGTACACGGAGCTGTTCCGGAT. Human *SCARB1*, forward, CGCGGTACCCCACCATGGGCTGCTCCGCCAAAGC; reverse, CAGCGCTGCGGCCGCCTACAGTTTTGCTTCCTGCA. Mouse *Scarb1*, forward, GCATACAGGGTACCGCCACCATGGGCGGCAGCTCCAGGGCGCGC; reverse, GCATACAGCTCGAGCTATAGCTTGGCTTCTTGCAGCAC. Human *CLDN1*, forward, CGCGGTACCCCACCATGGCCAACGCGGGGCTGCA; reverse, ACGCTCGAGTCACACGTAGTCTTTCCCGC. Mouse *Cldn1*, forward, GCATACAGGGTACCGCCACCATGGCCAACGCGGGGCTGCAGCTG; reverse, GCATACAGCTCGAGTCACACATAGTCTTTCCCACTAGA. Human *OCLN*, forward, CGCGGTACCCCACCATGTCATCCAGGCCTCTTGA; reverse, ACGCTCGAGCTATGTTTTCTGTCTATCAT. Mouse *Ocln*, forward, GCATACAGGGTACCGCCACCATGTCCGTGAGGCCTTTTGAAAGT; reverse, GCATACAGCTCGAGCTAAGGTTTCCGTCTGTCATAATC.

Production of recombinant adenoviruses. Adenoviral stocks were generated as previously described[30]. Briefly, adenoviral constructs were transfected into HEK293 cells (American Type Culture Collection) using the calcium-phosphate method. Transfected cultures were maintained until cells exhibited full cytopathic effect (CPE), then harvested and freeze-thawed. Supernatants were serially passaged two more times with harvest at full CPE and freeze thawed. For virus purification, cell pellets were resuspended in 0.01 M sodium phosphate buffer pH 7.2 and lysed in 5% sodium-deoxycholate, followed by DNase I digestion. Lysates were centrifuged and the supernatant was layered onto a 1.2–1.46 g ml^{-1} CsCl gradient, then spun at 95,389*g* on a Beckman Optima 100K-Ultra centrifuge using an SW28 spinning-bucket rotor (Beckman-Coulter). Adenovirus bands were isolated and further purified on a second CsCl gradient using an SW41.Ti spinning-bucket rotor. Resulting purified adenoviral bands were isolated using a 18.5G needle and twice-dialysed against 4% sucrose. Adenovirus concentrations were measured at 10^{12} times the OD_{260} reading on a FLUOstar Omega plate reader (BMG Labtech). Adenovirus stocks were aliquoted and stored at −80 °C.

Production of recombinant vaccinia virus. rVV expressing HCV-1 C-NS2 (rVV E12 C/B from Chiron Corporation) was obtained through the National Institutes of Health AIDS Research and Reference Reagent Program[31,32]. rVV was amplified by infecting HeLa S3 cells in suspension culture at a multiplicity of infection of 0.5 for 48 h. Intracellular virus was released by three freeze–thaw cycles and sonication. The virus was pelleted over a 36% sucrose cushion and resuspended in PBS[33].

HCV genome construction. Jc1FLAG2(p7Fluc2A) is a fully infectious HCV reporter genome similar to Jc1FLAG2(p7nsGluc2A)[34]. This monocistronic genome encodes a Flag epitope, followed by a Gly-Ser-Gly-Ala linker, fused to the N terminus of E2. The Fluc reporter, in tandem with the foot and mouth disease virus autoproteolytic peptide sequence (2A), was inserted between p7 and NS2. Bi-nlsCre-JC1FLAG2 (HCV-CRE) was created by amplification of nlsCre-recombinase from a TRIP-based construct, followed by insertion into the MluI/PmeI sites of a wild-type or polymerase defective (GNN) BiYpet-JC1FLAG2 (ref. 14) genome, replacing Ypet. nlsCre was similarly inserted into bicistronic versions of intergenotypic chimaeras encoding the core-NS2 sequences of Con1 (1b), ED43 (4a), HK6a (6a) and QC69 (7a), which have been previously described[35]. All plasmid constructs were verified by DNA sequencing. Primer sequences for CRE were: forward, CCCAACGCGTATGCCCAAGAAGAAGAGGAAGGTGTCCA; reverse, AGGGTTTAAACTTACTTGTACAGATCGCCATCTTC.

HCV generation and infections. Huh-7.5.1 or Huh-7.5 cells were electroporated with *in vitro* transcribed full-length HCV RNA. Seventy-two hours after electroporation, the medium was replaced with DMEM containing 1.5% FBS and supernatants were harvested every six hours starting from 72 h. Pooled supernatants were clarified by centrifugation at 1,500*g*, filtered through a 0.45 μm bottle top filter (Millipore) and concentrated using a stirred cell (Millipore). Viral titres ($TCID_{50}$) were determined using Huh-7.5 cells as previously described[28].

E2 antibody ELISAs. To determine end-point titres (EPT) of mouse sera, microplates (Corning Costar 3690) were coated with *Galanthus nivalis* lectin (GNL; Sigma) at 5 μg ml^{-1} overnight at 4 °C. Microwells were washed four times with PBS containing 0.05% Tween 20 and blocked with non-fat milk (NFM, 4%; BioRad) diluted in the wash buffer. HCV glycoproteins E1 and E2 were produced by transient transfection of 293T cells with phCMV-H77 and solubilised in lysis buffer (25 mM Tris pH 7.6, 140 mM NaCl, 1% Triton X-100, 0.5% NP-40 & 0.02% sodium azide). To adsorb E1E2, transfected cell lysate was diluted 1:30 in wash buffer containing NFM (1%) and incubated in the microwells for 1.5 h at room temperature (22 °C). Following blocking and washing, serially diluted mouse sera were added to the ELISA plates and incubated at room temperature for 1 h. Microwells were washed four times and incubated with peroxidase-conjugated goat anti-mouse IgG F(ab′)$_2$ (diluted 1:2,000) at room temperature for 1 h. Following detection with tetramethylbenzidine substrate (TMB; Pierce), absorbance at 450 nm was measured with a micoplate reader (Molecular Devices). The EPT of each serum was defined as the reciprocal of the dilution giving a threefold higher signal than the negative control. The positive control was a mixture of sera from 12 mice immunized with E2 and diluted at 1:2,000; the negative control was serum from non-immunized Rosa26-Fluc mice. Owing to high non-specific reactivity in mouse sera, only an EPT $>$200 is considered a true positive signal in this assay. Each ELISA was done in duplicate.

Histological detection of HCV entry factors. Liver and spleen of mice injected with adenoviruses encoding human entry factors were harvested 24 h after injection and fixed using 4% paraformaldehyde. Tissue sections (8 μm) were deparaffinized and subjected to antigen retrieval by boiling for 30 min in citrate buffer (10 mM sodium citrate, 0.05% Tween 20, pH 6.0). Entry factors were stained with human-specific primary antibodies for 16 h at 4 °C followed by secondary antibody staining using Alexa 488 or Alexa 633-conjugated antibodies for 2 h at room temperature. For *in situ* detection of EGFP fluorescence, mouse tissue was immediately frozen in O.C.T. (Optimal Cutting Temperature) compound at −80 °C. Tissue sections (∼5–6 μm) were cut on poly-L-lysine coated slides. Secondary antibodies goat anti-mouse or goat anti-rabbit Alexa 488- or rhodamine conjugates (Invitrogen; 1:1,000) were used for immunofluorescence. Nuclei were detected using DAPI in VectaShield Mounting medium (Vector Laboratories). Images were captured on an Axioplan 2 imaging fluorescence microscope (Zeiss) using Metavue Software (Molecular Devices). Images were processed using ImageJ software (NIH).

Isolation of murine hepatocytes. Mice were anaesthetized by intraperitoneal injection of a mixture of 100 mg kg^{-1} ketamine and 10 mg kg^{-1} xylazine. Livers were perfused through the inferior vena cava for 5 min each with chelating buffer (0.5 mM EGTA, 0.05 M HEPES pH 7.3 in Ca/Mg-free HBSS) at a flow rate of 2 ml min^{-1} followed by collagenase solution (4.8 mM $CaCl_2$, 100 U ml^{-1} collagenase type IV,

0.05 M HEPES pH 7.3 in Ca/Mg-free HBSS). The resulting cell suspension was passed through a 100 μm cell strainer, washed twice in HBSS and was fixed in 4% paraformaldehyde. Purity of isolated hepatocytes was over 90% in all preparations as confirmed by intracellular staining for murine albumin.

Western blotting. Perfused murine liver tissue was homogenized in lysis buffer containing 1% Triton X-100, 50 mM Tris-HCl pH 8, 150 mM NaCl, and Mini EDTA-free Protease Inhibitor Cocktail (Roche) for 30 min on ice. Fifteen micrograms of protein lysate was separated on 4–12% Bis/Tris NuPage polyacrylamide gels (Invitrogen). Proteins were transferred to nitrocellulose membranes and entry factors were detected using antibodies against CD81 (1:200), human SCARB1 (1:500), CLDN1 (1:200) or OCLN (1:200). β-Actin (1:10,000) was probed as a loading control. Following secondary antibody staining with HRP-conjugated anti-mouse IgG Fc (JIR, 1:10,000), western blots were visualized using SuperSignal West Pico (Thermo Scientific). 786O, 293T and HepG2 cell lysates, deficient in OCLN, CLDN1 and CD81, respectively, were used as negative controls. Huh-7.5 cell lysates served as positive control.

RT–PCR quantification of cellular RNAs. To quantify expression of human and murine genes (entry factors and interferon-stimulated genes, ISGs), the livers of FVB/NJ mice were harvested 24 h after adenovirus injection. Total liver RNA was isolated using RNeasy isolation kit (Qiagen) and cDNA was synthesized from 0.5 μg RNA using a SuperScript VILO cDNA Synthesis Kit (Invitrogen) according to manufacturer's instructions. Quantitative PCR was performed with a light cycler LC480 (Roche Applied Science) using an Applied Biosystems SYBR Green PCR Master Mix and the following primer pairs. Human *CD81*, forward, TGTTCTTGAGCACTGAGGTGGTC; reverse, TGGTGGATGATGACGCCAAC. Human *SCARB1*, forward, CGGATTTGGCAGATGACAGG; reverse, GGGGGAGACTCTTCACACATTCTAC. Human *CLDN1*, forward, CACCTCATCGTCTTCCAAGCAC; reverse, CCTGGGAGTGATAGCAATCTTTG. Human *OCLN*, forward, CGGCAATGAAACAAAAGGCAG; reverse, GGCTATGGTTATGGCTATGGCTAC. Mouse *Cd81*, forward, GGCTGTTCCTCAGTATGGTGGTAG; reverse, CCAAGGCTGTGGTGAAGACTTTC. Mouse *Scarb1*, forward, CAAAAAGCATTTCTCCTGGCTG; reverse, AATCTGTCAAGGGCATCGGG. Mouse *Cldn1*, forward, TTATGCCCCCAATGACAGCC; reverse, ATGAGGTGCCTGGAAGATGATG. Mouse *Ocln*, forward, ACTAAGGAAGCGATGAAGCAGAAG; reverse, GCTCTTTGGAGGAAGCCTAAACTAC. Mouse *Gapdh*, forward, ACGGCCGCATCTTCTTGTGCA; reverse, ACGGCCAAATCCGTTCACACC. Mouse viperin, forward, TGCTGGCTGAGAATAGCATTAGG; reverse, GCTGAGTGCTGTTCCCATCT. Mouse *Ifi27l2a*, forward, GCTTGTTGGGAACCCTGTTTG; reverse, GGATGGCATTTGTTGATGTGGAG. Mouse *Ifi44*, forward, AACTGACTGCTCGCAATAATGT; reverse, GTAACACAGCAATGCCTCTTGT. Mouse *Mx1*, forward, GACCATAGGGGTCTTGACCAA; reverse, AGACTTGCTCTTTCTGAAAAGCC. Mouse *Eif2ak2*, forward, ATGCACGGAGTAGCCATTACG; reverse, TGACAATCCACCTTGTTTTCGT. Mouse *Oas1*, forward, ATGGAGCACGGACTCAGGA; reverse, TCACACACGACATTGACGGC. Mouse *Ifnb1*, forward, CAGCTCCAAGAAAGGACGAAC; reverse, GGCAGTGTAACTCTTCTGCAT. Mouse *Cxcl10*, forward, CCAAGTGCTGCCGTCATTTTC; reverse, GGCTCGCAGGGATGATTTCAA.

RT–PCR quantification of HCV RNA. Total RNA was isolated from mouse brain, liver and sera using the RNAeasy kit (Qiagen). HCV genome copy number was quantified by one-step RT–PCR using Multicode-RTx HCV RNA Kit, (EraGen) and a light cycler LC480 (Roche Applied Science), according to manufacturers' instructions.

Bioluminescence imaging. Unless otherwise specified, mice were injected with 10^{11} adenoviral p.f.u. 24 h before intravenous injection with 2×10^7 $TCID_{50}$ HCV-CRE. At 72 h after infection, mice were anaesthetized using ketamin/xylazine and injected intraperitoneally with 1.5 mg luciferin (Caliper Lifesciences). Bioluminescence was measured using an IVIS Lumina II platform (Caliper Lifesciences).

***In vitro* neutralization assay.** Serum of FVB/NJ mice, either mock immunized or immunized with rVV-HCV1 were collected and pooled 5 weeks after immunizations. Serial dilutions of mouse serum or monoclonal anti-HCV E2 antibody (clone 3/11) were pre-incubated with intergenotypic JFH1 chimaeras expressing the structural proteins of genotypes 1a (H77), 1b (Con1 and J4), 2a (J6), 3a (S52), 4a (ED43), 5a (SA13), 6a (HK6a) or 7a (QC69) for 1 h at 4 °C. Supernatants were used to infect naive Huh7.5 cells at a calculated multiplicity of infection of 0.1 for 6 h after which cell were washed and medium was replaced. Cells were collected 48 h after infection and were stained for expression of HCV NS5a and were analysed by flow cytometry.

Flow cytometry. Immune activation, depletion efficiency as well as the frequency of infected hepatocytes were confirmed by flow cytometry using an LSRII flow cytometer (BD Biosciences). For immune activation and depletion efficiency, peripheral blood mononuclear cells (PBMCs) and splenocytes of mice were isolated and purified via density gradient centrifugation. Cells were stained with directly fluorochrome-conjugated antibodies directed against CD3, CD4, CD8, CD11b and CD49b. For the determination of infection frequency, hepatocytes were isolated from Rosa26-GNZ mice, fixed in 4% paraformaldehyde, permeabilized in PBS + 0.01% Triton X-100 and stained with fluorochrome-conjugated antibodies against murine albumin and CD81. Data were analysed using Flowjo software (Treestar Software).

***In vivo* depletion.** Mice were injected intraperitoneally with 100 mg kg^{-1} of either anti-mouse Ly-6G, anti-mouse Asialo GM1 or a mixture of anti-mouse CD4 and CD8. Depletion of the cell populations was verified 24 h after injection by flow cytometry. Adenoviral injection of HCV entry factor constructs was initiated following confirmation of depletion.

Statistical analysis. Statistical analyses were performed using Graphpad Prism Software. Statistics were calculated using Kruskal–Wallis one-way analysis of variance. *P* values below 0.05 were considered statistically significant.

28. Lindenbach, B. D. *et al.* Complete replication of hepatitis C virus in cell culture. *Science* **309,** 623–626 (2005).
29. Flint, M. *et al.* Characterization of hepatitis C virus E2 glycoprotein interaction with a putative cellular receptor, CD81. *J. Virol.* **73,** 6235–6244 (1999).
30. Schoggins, J. W., Gall, J. G. & Falck-Pedersen, E. Subgroup B and F fiber chimeras eliminate normal adenovirus type 5 vector transduction *in vitro* and *in vivo*. *J. Virol.* **77,** 1039–1048 (2003).
31. Selby, M. *et al.* Hepatitis C virus envelope glycoprotein E1 originates in the endoplasmic reticulum and requires cytoplasmic processing for presentation by class I MHC molecules. *J. Immunol.* **162,** 669–676 (1999).
32. Cooper, S. *et al.* Analysis of a successful immune response against hepatitis C virus. *Immunity* **10,** 439–449 (1999).
33. Law, M. & Smith, G. L. in *Vaccinia Virus and Poxvirology Methods and Protocols Methods in Molecular Biology Series* (ed. Isaacs, S. N.) 187–204 (Humana, 2004).
34. Marukian, S. *et al.* Cell culture-produced hepatitis C virus does not infect peripheral blood mononuclear cells. *Hepatology* **48,** 1843–1850 (2008).
35. Gottwein, J. M. *et al.* Development and characterization of hepatitis C virus genotype 1–7 cell culture systems: role of CD81 and scavenger receptor class B type I and effect of antiviral drugs. *Hepatology* **49,** 364–377 (2009).

LETTER

doi:10.1038/nature10135

Immunogenicity of induced pluripotent stem cells

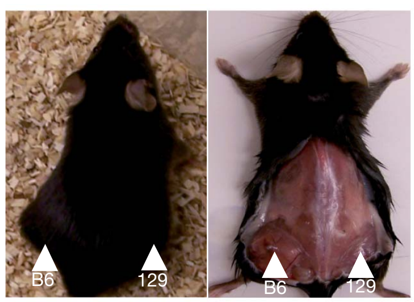
Tongbiao Zhao[1], Zhen-Ning Zhang[1], Zhili Rong[1] & Yang Xu[1]

Induced pluripotent stem cells (iPSCs), reprogrammed from somatic cells with defined factors, hold great promise for regenerative medicine as the renewable source of autologous cells[1–5]. Whereas it has been generally assumed that these autologous cells should be immune-tolerated by the recipient from whom the iPSCs are derived, their immunogenicity has not been vigorously examined. We show here that, whereas embryonic stem cells (ESCs) derived from inbred C57BL/6 (B6) mice can efficiently form teratomas in B6 mice without any evident immune rejection, the allogeneic ESCs from 129/SvJ mice fail to form teratomas in B6 mice due to rapid rejection by recipients. B6 mouse embryonic fibroblasts (MEFs) were reprogrammed into iPSCs by either retroviral approach (ViPSCs) or a novel episomal approach (EiPSCs) that causes no genomic integration. In contrast to B6 ESCs, teratomas formed by B6 ViPSCs were mostly immune-rejected by B6 recipients. In addition, the majority of teratomas formed by B6 EiPSCs were immunogenic in B6 mice with T cell infiltration, and apparent tissue damage and regression were observed in a small fraction of teratomas. Global gene expression analysis of teratomas formed by B6 ESCs and EiPSCs revealed a number of genes frequently overexpressed in teratomas derived from EiPSCs, and several such gene products were shown to contribute directly to the immunogenicity of the B6 EiPSC-derived cells in B6 mice. These findings indicate that, in contrast to derivatives of ESCs, abnormal gene expression in some cells differentiated from iPSCs can induce T-cell-dependent immune response in syngeneic recipients. Therefore, the immunogenicity of therapeutically valuable cells derived from patient-specific iPSCs should be evaluated before any clinical application of these autologous cells into the patients.

To vigorously examine the immunogenicity of cells derived from iPSCs, we took advantage of the capability of ESCs and iPSCs to form teratomas in mice that allows the simultaneous evaluation of the immunogenicity of various cell types derived from them. Whereas B6 ESCs could efficiently form teratomas in B6 mice without any evidence of immune rejection as indicated by the lack of any detectable $CD4^+$ T cell infiltration, a hallmark of immune rejection, the allogeneic 129/SvJ (129) ESCs were rapidly rejected before forming detectable teratomas in the same B6 recipients with massive infiltration of $CD4^+$ T cells into one detectable teratomas formed by 129 ESCs (Fig. 1a–d, Supplementary Fig. 1). The $CD4^+$ cells were not directly differentiated from the implanted ESCs because no $CD4^+$ cells were detectable in any examined teratomas formed by B6 and 129 ESCs in severe combined immunodeficient (SCID) mice (Fig. 1d). B6 and 129 ESCs had similar proliferation rates and both could efficiently form teratomas in SCID mice (Supplementary Fig. 1a–e). Therefore, these findings validate the feasibility of using this teratoma formation assay to evaluate the immunogenicity of iPSC derivatives *in vivo*.

We initially established ViPSCs from B6 MEFs with the cocktails of retrovirus expressing either three (Oct4/Sox2/Klf4) or four (Oct4/Sox2/myc/Klf4) reprogramming factors as described[1]. The subcloned ViPSCs had normal karyotypes, expressed ESC-specific surface markers and pluripotency genes, and were pluripotent as indicated by their capability to form teratomas in SCID mice and contribute to adult chimaeric mice (Supplementary Fig. 2a–g). Four independent iPSC clones, two reprogrammed with three factors (V3-1 and V3-3) and two with four factors (V4-1 and V4-2), were selected for further analysis (Supplementary Fig. 2h). Most implanted B6 ViPSCs failed to form detectable teratomas or formed teratomas that were subsequently immune-rejected with T cell infiltration and massive necrosis (Supplementary Fig. 3a–e). The teratomas that did not undergo apparent regression were also infiltrated with $CD4^+$ T cells with apparent

a

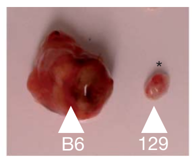


b

Cell line (number)	Day 21 (%)	Day 30 (%)
C57BL/6 (1 x 10^6)	12/12 (100)	18/19 (95)
(3 x 10^6)	–	12/12 (100)
129 (1 x 10^6)	0/12 (0)	0/19 (0)
(3 x 10^6)	–	1*/12 (8.3)

– Not determined

c



d

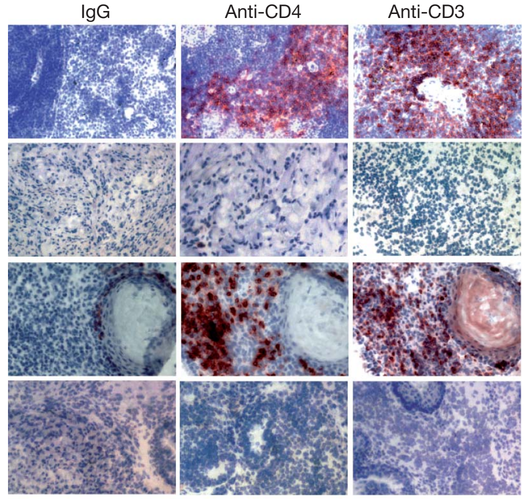


Figure 1 | Immunogenicity of syngeneic and allogenic ESCs in male B6 mice. a, B6 but not 129 ESCs can efficiently form teratomas in B6 mice after subcutaneous injection. The teratomas shown is 30 days after implantation. **b**, Summary of teratomas formation by ESCs in B6 mice 21 and 30 days after implantation. Only one small teratoma formed by 129 ESCs was detected (asterisk) and is shown in **c**. **d**, Infiltration of T cells was detected in the teratomas formed by 129 ESCs but not the ones formed by B6 ESCs in B6 mice. T cells were identified by anti-CD4 and anti-CD3 antibodies. Sections from the spleen and teratomas formed by 129 ESCs in SCID mice were used as positive and negative controls.

[1]Section of Molecular Biology, Division of Biological Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0322, USA.

necrosis within parts of the tumour (Supplementary Fig. 3d, e). Therefore, cells derived from B6 ViPSCs are highly immunogenic in B6 mice.

Recent studies have shown the existence of T cells specific for the cells expressing Oct4 in the periphery[6]. Therefore, the reactivation of Oct4 expression in cells differentiated from B6 ViPSCs could induce immune responses in B6 mice (Supplementary Fig. 2i). To address this issue, we developed a novel episomal approach to reprogram B6 MEFs into EiPSCs that express ESC markers and pluripotency genes as well as contribute to adult chimaeric mice (Fig. 2a–e). Extensive Southern blotting analysis demonstrated that some EiPSC clones (1E12, 1E13, 3E1) had lost the episomal vector and harboured no random integration of the reprogramming vector (Fig. 2f). The expression cassette was excised from the genome of 2E2 iPSC clone that harboured one random integration of the episomal vector by transient expression of Cre enzyme (Supplementary Fig. 4).

EiPSCs had normal karyotypes and efficiently formed teratomas in B6 mice. However, the majority of teratomas derived from EiPSCs of both early and late passages showed apparent infiltration of T cells (Figs 3a, d and Supplementary Fig. 5). In addition, apparent tumour regression with extensive tissue necrosis was detected in 10% of teratomas formed by EiPSCs in B6 mice within 2 months of implantation (Fig. 3b, c). No apparent tumour regression was observed in the majority of the teratomas formed by EiPSCs in B6 mice before they reached the allowed maximal size (Fig. 3c). Therefore, we concluded that cells derived from B6 EiPSCs can be immunogenic in B6 recipients, but their overall immunogenicity is lower than the cells derived from B6 ViPSCs.

To determine the generality of our conclusion, two independently generated integration-free B6 iPSC lines, which were reprogrammed from B6 MEFs with a plasmid vector expressing Oct4/Sox2/Myc/Klf4 (ref. 7), were implanted into B6 mice. T cell infiltration was observed in most teratomas formed by these B6 iPSCs in B6 mice, some of which also exhibit tissue necrosis (Supplementary Fig. 6). In addition, a small fraction of teratomas had undergone apparent regression by 40 days after implantation. These findings support the conclusion that cells derived from iPSCs are immunogenic in syngeneic recipients.

To understand the basis of this immunogenicity, the profile of gene expression in teratomas derived from B6 ESCs and EiPSCs revealed a number of genes overexpressed in teratomas derived from B6 EiPSCs (Supplementary Fig. 7a). Expression analysis of six regressing teratomas formed by two independent B6 EiPSCs in B6 mice indicated that 9 of the 23 tested genes (*Lce1f*, *Spt1*, *Cyp3a11*, *Zg16*, *Lce3a*, *Chi3L4*, *Olr1*, *Retn*, *Hormad1*) were commonly overexpressed in these teratomas (Fig. 4a). *Hormad1* has been identified as a tumour antigen and *Spt1* as a tissue-specific antigen[8,9].

To test the possibility that the abnormal expression of these genes in teratomas derived from B6 iPSCs contributes to their immunogenicity

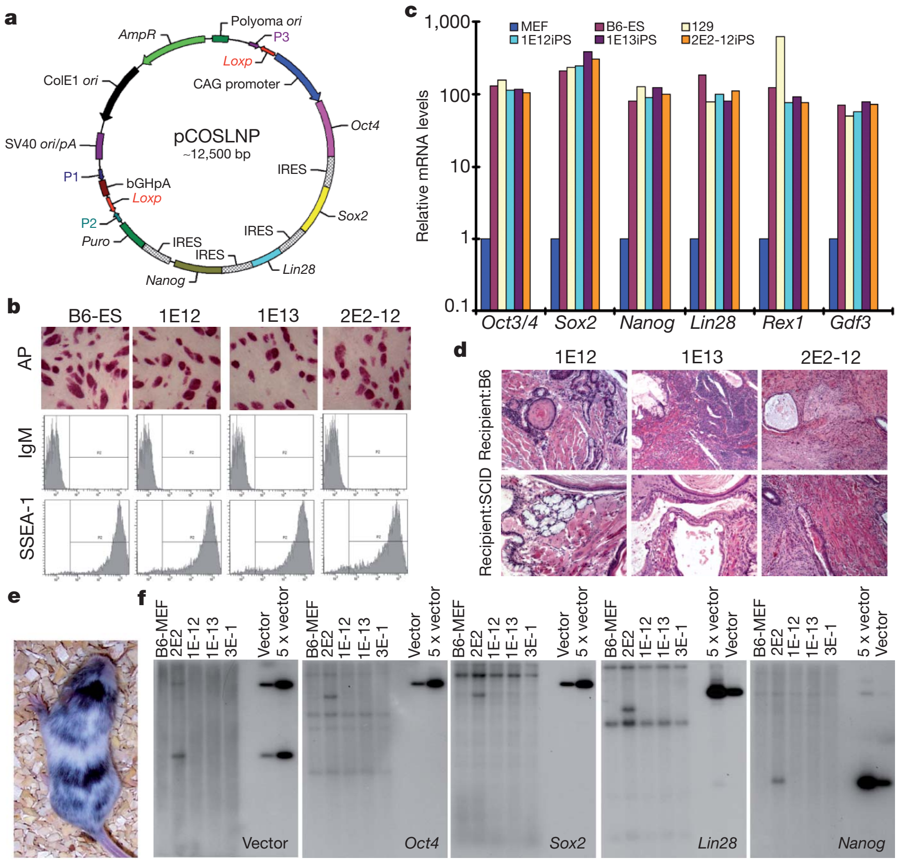
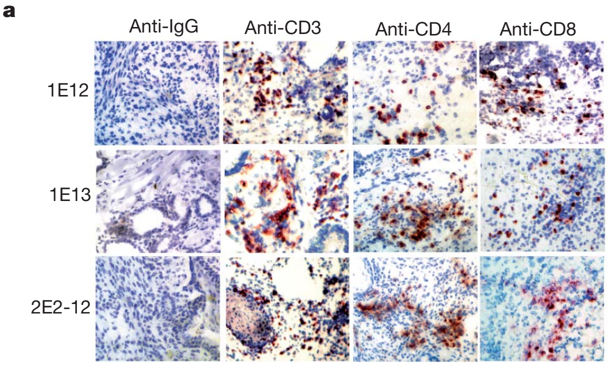
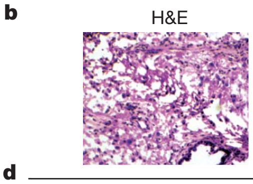


Figure 2 | A new episomal approach to generate EiPSCs from B6 MEFs. **a**, Diagram of the episomal vector that expresses the four reprogramming factors (Oct4/Sox2/Nanog/Lin28) and puromycin resistance gene from one messenger RNA separated by IRES sequences. The entire expression cassette is flanked with LoxP sites. **b**, **c**, EiPSCs were positive for alkaline phosphatase (AP) and SSEA-1 (**b**) and expressed pluripotency genes to the same levels as those of B6 ESCs as determined by quantitative real-time PCR (**c**). The mRNA levels in MEFs are arbitrarily set to 1. **d**, EiPSCs form teratomas in both B6 mice (top panel) and SCID mice (bottom panel). **e**, EiPSCs can contribute to adult chimaeric mice after injecting into the blastocysts derived from albino mice. **f**, Southern blotting analysis indicates no random integration of the episomal vectors in EiPSC clones 1E-12, 1E-13 and 3E-1. Clone 2E2 has one copy of the episomal vector integrated into the genome. Genomic DNA derived from iPSCs was digested with BamHI and hybridized to various probes that together cover the entire episomal vector.





c

	Number	Ratio (%)
No tumour	7/64	11
Regressing tumour	6/64	9
Growing tumour	51/64	80

d

EiPSCs	Day 30	CD4 infiltration	Day 42	CD4 infiltration	Day 56	CD4 infiltration
1E12	11/13 (85)	3/6 (50)	6/7 (86)	2/3 (67)	4/5 (80)	3/4 (75)
1E13	10/11 (91)	3/4 (75)	5/7 (71)	2/3 (67)	3/4 (75)	3/3 (100)
2E2-12	5/7 (71)	2/3 (67)	4/5 (80)	2/2 (100)	4/5 (80)	3/3 (100)

Figure 3 | Cells derived from B6 EiPSCs can be immunogenic in B6 mice. **a**, T-cell infiltration was detected in the majority of teratomas formed by B6 EiPSCs in male B6 mice. 2E2-12 iPSCs is a subclone of 2E2 clone after LoxP/Cre-mediated deletion of the reprogramming factor expression cassette from the integrated copy of episomal vector. **b**, Tissue necrosis was detected in the regressing teratomas formed by B6 EiPSCs in male B6 mice. H&E, haematoxylin and eosin staining. **c**, Summary of teratoma formation by B6 EiPSCs in male B6 mice. **d**, Summary of teratoma formation and $CD4^+$ T cell infiltration at different time points after implantation of EiPSCs in male B6 mice.

in B6 mice, seven such genes were ectopically expressed in B6 ESCs and their derived teratomas under the control of the ubiquitously active CAG promoter/enhancer (Supplementary Fig. 7b). Like B6 ESCs, over 90% of implants of B6 ESCs with empty vector as well as transgenic *Lce1f*-B6 ESC and *Retn*-B6 ESCs formed teratomas in B6 mice (Fig. 4b). In contrast, over 80% of *Zg16*-B6 ESC implants and 50% of *Hormad1*- or *Cyp3a11*-B6 ESC implants failed to form visible teratomas in B6 mice (Fig. 4b). Extensive T cell infiltration and widespread necrosis were detected in the teratomas formed by *Zg16*- and *Hormad1*-B6 ESCs in B6 mice but rarely detectable in the teratomas derived from *Lce1f*- and *Retn*-B6 ESCs in B6 mice (Fig. 4c, d). To rule out the possibility that the regression of the teratomas formed by *Zg16*- and *Hormad1*-B6 ESCs in B6 mice is secondary to the abnormal proliferation or cell death induced by the ectopic expression of these genes, the proliferation and survival of *Zg16*- and *Hormad1*-B6 ESCs were identical to B6 ESCs (Supplementary Fig. 7c, e). In addition, the weight of the teratomas formed by *Zg16*- and *Hormad1*-B6 ESCs in SCID mice was similar to that of B6 ESCs (Supplementary Fig. 7d).

To identify the immune responses against the cells derived from iPSCs, we used $CD4^{-/-}$ and $CD8^{-/-}$ B6 mice to examine the importance of T cells in the immune rejection. The robust immune rejection of the teratomas formed by B6 ViPSCs as well as *Zg16*- and *Hormad1*-B6 ESCs in B6 mice was abolished in both $CD4^{-/-}$ and $CD8^{-/-}$ B6 mice

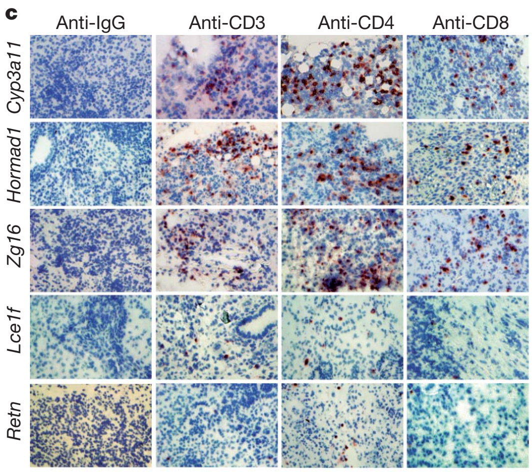
a

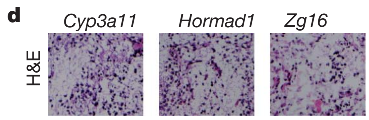
Gene name	Regressing tumour 1	2	3	4	5	6
Spt1 (NM009267)	+	+	+	+	+	+
Cyp3a11 (NM007818)	+	+	+	+	+	+
Lce3a (NM001039594)	+	+	+	+	+	+
Lce1f (NM026394)	+	+	+	+	+	+
Zg16 (NM026918)	+	+	+	+	+	+
Hormad1 (NM026489)	−	+	+	+	+	+
Retn (NM022984)	+	+	+	+	−	+
Chi3l4 (NM145126)	+	+	+	+	−	+
Olr1 (NM138648)	+	+	+	+	−	+

+ Overexpressed −Undetectable

b

Cell lines	Day 30 (%)
Vector-B6	17/18 (94.4)
Zg16-B6	1/8 (12.5)
Hormad1-B6	6/12 (50)
Cyp3a11-B6	6/12 (50)
Lce3a-B6	11/16 (68.8)
Spt1-B6	15/16 (93.8)
Lce1f-B6	8/8 (100)
Retn-B6	8/8 (100)





e

Cell lines	$CD4^{-/-}$(Day 30)	$CD8^{-/-}$(Day 30)
Zg16-B6	8/8 (100)	8/8 (100)
Hormad1-B6	7/8 (87.5)	8/8 (100)
1E12-iPS	18/20 (90)	20/20 (100)
V3-1-iPS	3/3 (100)	3/3 (100)

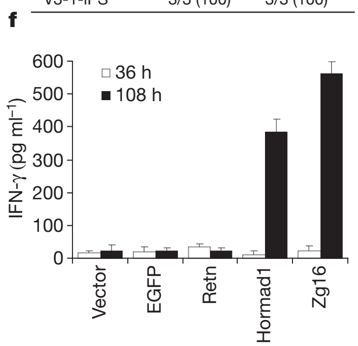


Figure 4 | Abnormal overexpression of some proteins contributes directly to the immunogenicity of cells derived from B6 EiPSC in B6 mice. **a**, Nine genes were found to be commonly overexpressed in six regressing teratomas formed by two B6 EiPSCs. The expression of 23 genes identified as overexpressed in EiPSC-derived teratomas by microarray analysis was analysed by real-time PCR. **b**, Summary of teratoma formation by various transgenic B6 ESCs in male B6 mice. **c**, Extensive infiltration of T cells in the teratomas formed by *Cyp3a11*-, *Hormad1*- and *Zg16*-B6 ESCs in B6 mice. Few infiltrating T cells were detectable in the teratomas formed by *Lce1f*- and *Retn*-B6 ESCs in B6 mice. Representative images are shown. **d**, Extensive necrosis is present in teratomas formed by *Cyp3a11*-, *Hormad1*- and *Zg16*-B6 ESCs in B6 mice. **e**, The immune rejection of the teratomas formed by *Hormad1*-B6 ESCs, *Zg16*-B6 ESCs, B6 EiPSCs and B6 ViPSCs is abolished in $CD4^{-/-}$ or $CD8^{-/-}$ B6 mice. **f**, IFN-γ release assay to detect the presence of primed T cells specific for cells expressing Hormad1 and Zg16 in B6 mice harbouring the teratomas formed by EiPSCs. Each data point represents the mean of duplicate cultures. Consistent data are obtained from three independent experiments.

(Fig. 4e). In addition, no regression of teratomas formed by EiPSCs in $CD4^{-/-}$ or $CD8^{-/-}$ B6 mice was observed. Therefore, both $CD4^{+}$ helper T cells and $CD8^{+}$ cytotoxic T cells are critical for this immune rejection. These findings also indicate that the innate immunity does not have an important role in the immune rejection of the cells derived from iPSCs.

To further determine whether the abnormal expression of Hormad1 and Zg16 in teratomas formed by EiPSCs directly activates T-cell responses in B6 mice, we performed the IFN-γ releasing assay that measures the antigen-specific activation of *in-vivo*-primed T cells[10]. Dendritic cells purified from B6 mice were transfected with either empty expression vector or vectors expressing Zg16, Hormad1, Retn or EGFP (enhanced green fluorescent protein). LPS-matured dendritic cells expressing Hormad1 or Zg16 but not the dendritic cells expressing Retn or EGFP could induce IFN-γ production from purified T cells, indicating the presence of primed T cells specific for cells expressing Hormad1 or Zg16 in B6 mice harbouring the teratomas formed by EiPSCs (Fig. 4f). Although these findings did not identify the specific peptides responsible for activating T cells, they demonstrate that the abnormal expression of Hormad1 and Zg16 contributes directly to the immunogenicity of the cells derived from EiPSCs in syngeneic recipients. Hormad1 was also overexpressed in most teratomas formed by four independently generated integration-free iPSCs reprogrammed with adenoviral vectors, recombinant proteins or plasmid vectors[7,11,12]. In addition, Zg16 was overexpressed in most teratomas formed by iPSCs reprogrammed with recombinant proteins. Therefore, the abnormal expression of such immunogenic proteins could represent a common mechanism to induce T cell-mediated immune responses to cells derived from iPSCs.

Our findings indicate that some cells derived from iPSCs can be immunogenic in syngeneic recipients. The T-dependent immune response is likely due to the abnormal expression of antigens not expressed during normal development or differentiation of ESCs, leading to the break of peripheral tolerance. The expression of these minor antigens could be due to the subtle yet apparent epigenetic difference between iPSCs and ESCs[13–20]. In addition, recently discovered mutations in the coding sequences of iPSCs could also contribute to the immunogenicity of iPSC derivatives[21]. Therefore, for the clinical development of iPSCs, current reprogramming technology needs to be optimized to minimize the epigenetic difference between iPSCs and ESCs. The *in vivo* immunogenicity test described here can provide a robust screening platform for improving the reprogramming technology.

METHODS SUMMARY

Mice. B6 mice and ESCs were purchased from The Jackson Laboratory. Only male mice were used in the transplantation studies. All animal experiments were performed in accordance with relevant guidelines and regulations, and approved by the Institutional Animal Care and Use Committee (IACUC).

iPSC generation and characterization. MEFs were isolated from B6 embryo as described[22]. For ViPSC production, MEFs were transduced with retrovirus cocktail as described[1]. For EiPSCs generation, MEFs were transfected with the episomal vector expressing the reprogramming factors. The lack of random integration of the episomal vector was confirmed by Southern blotting analysis with a combination of probes that cover the entire episomal vector.

Interferon-γ releasing assay. Dendritic cells from B6 mice were isolated and transfected with expression vectors. The transfected dendritic cells were matured by lipopolysaccharide (LPS) treatment for 12 h. T cells were purified from the pooled spleens and lymph nodes of five B6 mice harbouring the teratomas formed by EiPSCs and co-cultured with LPS-matured dendritic cells. Supernatant were collected at indicated time point to determine the IFN-γ levels.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 July 2010; accepted 19 April 2011.
Published online 13 May 2011.

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
2. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
3. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318,** 1917–1920 (2007).
4. Park, I.-H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451,** 141–146 (2008).
5. Lowry, W. E. *et al.* Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc. Natl Acad. Sci. USA* **105,** 2883–2888 (2008).
6. Dhodapkar, K. M. *et al.* Natural immunity to pluripotency antigen OCT4 in humans. *Proc. Natl Acad. Sci. USA* **107,** 8718–8723 (2010).
7. Jincho, Y. *et al.* Generation of genome integration-free induced pluripotent stem cells from fibroblasts of C57BL/6 mice without c-Myc transduction. *J. Biol. Chem.* **285,** 26384–26389 (2010).
8. Chen, Y. T. *et al.* Identification of CT46/HORMAD1, an immunogenic cancer/testis antigen encoding a putative meiosis-related protein. *Cancer Immun.* **5,** 9 (2005).
9. Kont, V. *et al.* Modulation of Aire regulates the expression of tissue-restricted antigens. *Mol. Immunol.* **45,** 25–33 (2008).
10. Guermonprez, P., Valladeau, J., Zitvogel, L., Théry, C. & Amigorena, S. Antigen presentation and T cell stimulation by dendritic cells. *Annu. Rev. Immunol.* **20,** 621–667 (2002).
11. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced pluripotent stem cells generated without viral integration. *Science* **322,** 945–949 (2008).
12. Zhou, H. *et al.* Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell* **4,** 381–384 (2009).
13. Chin, M. H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5,** 111–123 (2009).
14. Liu, L. *et al.* Activation of the imprinted *Dlk1-Dio3* region correlates with pluripotency levels of mouse stem cells. *J. Biol. Chem.* **285,** 19483–19490 (2010).
15. Stadtfeld, M. *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465,** 175–181 (2010).
16. Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genet.* **41,** 1350–1353 (2009).
17. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467,** 285–290 (2010).
18. Polo, J. M. *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature Biotechnol.* **28,** 848–855 (2010).
19. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).
20. Zhao, T. & Xu, Y. p53 and stem cells: new developments and new concerns. *Trends Cell Biol.* **20,** 170–175 (2010).
21. Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471,** 63–67 (2011).
22. Song, H., Hollstein, M. & Xu, Y. p53 gain-of-function cancer mutants induce genetic instability by inactivating ATM. *Nature Cell Biol.* **9,** 573–580 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Abe, S. Ding and K. Hochedlinger for their generous supply of integration-free mouse iPSCs. We thank N. Shastri for his advice on how to identify antigen-specific T cells. We thank J. Fink and Blue Lake of Xu laboratory as well as UCSD Cancer Center pathologic core for technical support. This work was supported by a NIH grant and an Early Translational Award from California Institute for Regenerative Medicine to Y.X. (ET-01277).

Author Contributions T.Z. and Y.X. designed the experiments, analysed the data and wrote the manuscript. T.Z., Z.-N.Z. and Z.R. executed the experiments under the overall coordination of T.Z.

Author Information The microarray data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE28573. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.X. (yangxu@ucsd.edu).

METHODS

Mice, cell culture and reprogramming episomal vector construction. C57BL/6 (B6) inbred mouse strain and ESCs were purchased from The Jackson Laboratory. Only male mice were used in the transplantation studies of ESCs and iPSCs. All animal experiments were performed in accordance with relevant guidelines and regulations, and approved by the Institutional Animal Care and Use Committee (IACUC). The ESCs and iPSCs were grown on the feeder layer derived from B6 MEFs under standard conditions. The full-length cDNA of *Oct4*, *Sox2*, *Lin28* and *Nanog* was sequentially inserted downstream of the CAG promoter in the episomal vector, separated by IRES sequences (Fig. 2a). The fifth gene, the puromycin resistance gene, is at the 3′ end of this mRNA transcript, separated from *Nanog* cDNA by the IRES sequence. This episomal vector is denoted pCOSLNP (CAG–*Oct4*–*Sox2*–*Lin28*–*Nanog*–*Puro*). Two LoxP sites in the same orientation were inserted into the episomal vector flanking the entire expression cassette.

iPSC generation and characterization. MEFs were isolated from B6 embryo as previously described[22]. For ViPSC production, MEFs were transduced with retrovirus cocktail expressing Oct4, Sox2, Klf4 with or without c-Myc. The iPSC colonies were picked 18 days after infection as described[1]. For EiPSCs generation, MEFs were transfected with pCOSLNP vector using Basic Nucleofector Kit for Primary Mammalian Fibroblasts (Lonza) followed by puromycin selection for 3 days, and then plated on irradiated B6 MEF feeders. Three weeks later, the culture was replated on fresh feeder cells. iPSC colonies were picked 10 to 30 days after replating, and the lack of random integration of the episomal vector was confirmed by Southern blotting analysis with a combination of probes that cover the entire episomal vector.

Quantitative real-time PCR analysis. Total RNA was purified from fibroblasts, ES cells, iPS cells and teratomas with a RNeasy total RNA isolation kit (Qiagen). Total RNA (1 μg) was reversely transcribed into cDNA, which was analysed by quantitative real time PCR analysis as previously described[22]. The primers used were as follows: Oct4F, 5′-GGCTCTCCCATGCATTCAA-3′; Oct4R, 5′-TTTAACCCCAAAGCTCCAGG-3′; Sox2F, 5′-AAATCTCCGCAGCGAAACG-3′; Sox2R, 5′-CCCCAAAAAGAAGTCCCAAGA-3′; Lin28F, 5′-CTGCTGTAGCGTGATGGTTGA-3′; Lin28R, 5′-CCACCCAATGTGTTCTATTGCA-3′; NanogF, 5′-TCGCCATCACACTGACATGA-3′; NanogR, 5′-TGTGCAGAGCATCTCAGTAGCA-3′; Rex1F, 5′-ACGAGTGGCAGTTTCTTCTTGGGA-3′; Rex1R, 5′-TATGACTCACTTCCAGGGGGCACT-3′; Gdf3F, 5′-GATTGCTTTTTCTGCGGTCTGT-3′; Gdf3R, 5′-CCAAGTTCTTCAGTCGGTTGCT-3′. Primers used for detection of reprogramming factor deletion were as follows: Oct4F (43–63), 5′-CCTTCCTTCCCCATGGCGGGA-3′; IRESR1 (53–31), 5′-TTATTCCAAGCGGCTTCGGCCAG-3′; Sox2F (1292–1310), 5′-CCCCAGCAGACTTCACATGT-3′; IRES-R (221–202), 5′-AGGAACTGCTTCCTTCACGA-3′; IRESF2 (476–498), 5′-TCGGTGCACATGCTTTACATGTG-3′; Lin28R (369–352), 5′-CCGGAACCCTTCCATGTG-3′; NanogRTGA (1131–1111), 5′-TCACACGTCTTCAGGTTGCAT-3′; P1, 5′-CGCCATCTTCTGAAGCTGAATC-3′; P2, 5′-ACCGAAAGGAGCGCACGACCCCAT-3′; P3, 5′-CCTACTCAGACAATGCGATGCA-3′; GAPDHF, 5′-CCAGTATGACTCCACTCACG-3′; GAPDHR, 5′-GACTCCACGACATACTCAGC-3′; Lce1fF, 5′-CTGTAGCCTGGGTTCTGG-3′; Lce1fR, 5′-GACGATGGCGACGAAGAG-3′; Spt1F, 5′-TGAAACTCAGGCAGATAG-3′; Spt1R, 5′-TGTCAACGCCACTGTTCT-3′; Olr1F, 5′-TGGTGGTTCCCTGCTGCTA-3′; Olr1R, 5′-ATCCTGCTGAGTAAGGTTCG-3′; Zg16F, 5′-CATCACCGCCTTCCGTAT-3′; Zg16R, 5′-CGTTGAAACTTGTGCCTGA-3′; RetnF, 5′-TCCTTGTCCCTGAACTGC-3′; RetnR, 5′-ACGAATGTCCCACGAGCC-3′; Hormad1F, 5′-CCAGATTACCAACCACCAG-3′; Hormad1R, 5′-TGAAAAGGTGTTGGGACT-3′; Lce3aF, 5′-GGCAGTGGTCAGCAGTCT-3′; Lce3aR, 5′-TTGGGAAATCCATTAGAAGA-3′; Cyp3a11F, 5′-ATCCCATTGCTAATAGAC-3′; Cyp3a11R, 5′-ATCATCACTGTTGACCCT-3′; Chi3l4F, 5′-ATGGCTACACTGGAGAAA-3′; Chi3l4R, 5′-TGCTGGAAATCCCACAAT-3′.

Southern blotting analysis. Genomic DNA (10 μg) was digested with BamHI, separated on 1% agarose gel and transferred to a nylon membrane. For the analysis of ViPSCs, the membrane was hybridized to the *Oct4* cDNA probe. For the analysis of EiPSCs, the membrane was hybridized to the cDNA probe of *Oct4*, *Sox2*, *Lin28* and *Nanog* as well as the vector backbone probe.

Teratoma formation and immunohistochemistry analysis. ESCs or iPSCs were collected, washed twice with PBS, and injected subcutaneously into the hind leg region of B6 or SCID mice. One or three million cells were used for each injection. Tumours were measured and surgically removed from the euthanized mice at the indicated time point. Teratomas were fixed either with 4% formaldehyde or frozen in optimal cutting temperature (OCT) compound. Sections were stained with haematoxylin and eosin or with various antibodies such as IgG, anti-CD4, anti-CD3 (BD Biosciences) as we described previously[23].

Microarray assay. Total RNA was purified from the teratomas with an RNeasy total RNA isolation kit (Qiagen). Microarray assay was performed by SeqWright using an Affymetrix Mouse 430A 2.0 chip.

Flow cytometric analysis. About 5×10^5 ESCs or iPSCs were stained for the expression of ESC-specific surface marker with anti-SSEA-1 antibody (Stemgents). Isotype-matched normal antibodies were used as negative controls. The stained cells were analysed by a BD LSR-II using FACSDiva software (Becton Dickinson) as we previously described[22].

Interferon-γ releasing assay. To obtain dendritic cells from B6 mice with the dendritic cell purification kit (Miltenyi Biotec), bone marrow cells were isolated from B6 mice and grown in Petri dish at a density of 10^6 cells ml^{-1} in complete medium supplemented with 10 ng ml^{-1} granulocyte/macrophage colony-stimulating factor (GM-CSF) and 5 ng ml^{-1} IL-4 according to the manufacturer's recommendation. On day 9, dendritic cells were purified and transfected with expression vectors using a mouse dendritic cell Nucleofector kit according to the manufacturer's instruction (Lonza). The transfected dendritic cells were matured by LPS treatment for 12 h. T cells were purified from the pooled spleens and lymph nodes of five B6 mice harbouring the teratomas formed by EiPSCs through negative selection using a pan T cell isolation kit (Miltenyi Biotec). Purified T cells (10^6) were immediately co-cultured with LPS-matured dendritic cells (2×10^5). Supernatant were collected at indicated time point to determine the IFN-γ levels using an ELISA kit (Thermo Scientific).

23. Song, H., Chung, S.-K. & Xu, Y. Modeling Disease in Human ESCs Using an Efficient BAC-Based Homologous Recombination System. *Cell Stem Cell* **6,** 80–89 (2010).

LETTER

doi:10.1038/nature10160

In vivo imaging of T_{reg} cells providing immune privilege to the haematopoietic stem-cell niche

Joji Fujisaki[1,2], Juwell Wu[1,2,3], Alicia L. Carlson[1], Lev Silberstein[2,4], Prabhakar Putheti[5], Rafael Larocca[5], Wenda Gao[5], Toshiki I. Saito[6]†, Cristina Lo Celso[2,4], Hitoshi Tsuyuzaki[7], Tatsuyuki Sato[7], Daniel Côté[8], Megan Sykes[6]†, Terry B. Strom[5], David T. Scadden[2,4] & Charles P. Lin[1,2]

Stem cells reside in a specialized regulatory microenvironment or niche[1,2], where they receive appropriate support for maintaining self-renewal and multi-lineage differentiation capacity[1–3]. The niche may also protect stem cells from environmental insults[3] including cytotoxic chemotherapy and perhaps pathogenic immunity[4]. The testis, hair follicle and placenta are all sites of residence for stem cells and are immune-suppressive environments, called immune-privileged sites, where multiple mechanisms cooperate to prevent immune attack, even enabling prolonged survival of foreign allografts without immunosuppression[4]. We sought to determine if somatic stem-cell niches more broadly are immune-privileged sites by examining the haematopoietic stem/progenitor cell (HSPC) niche[1,2,5–7] in the bone marrow, a site where immune reactivity exists[8,9]. We observed persistence of HSPCs from allogeneic donor mice (allo-HSPCs) in non-irradiated recipient mice for 30 days without immunosuppression with the same survival frequency compared to syngeneic HSPCs. These HSPCs were lost after the depletion of FoxP3 regulatory T (T_{reg}) cells. High-resolution *in vivo* imaging over time demonstrated marked co-localization of HSPCs with T_{reg} cells that accumulated on the endosteal surface in the calvarial and trabecular bone marrow. T_{reg} cells seem to participate in creating a localized zone where HSPCs reside and where T_{reg} cells are necessary for allo-HSPC persistence. In addition to processes supporting stem-cell function, the niche will provide a relative sanctuary from immune attack.

Despite rapid advances in our understanding of adult stem-cell biology, the immunological attributes of somatic stem-cell niches have remained largely unexplored. We hypothesized that the HSPC niche within the bone marrow[6,7] is an immune-privileged site based on the observation that human bone marrow contains a higher frequency of $CD4^+CD25^+FoxP3^+$ regulatory T (T_{reg}) cells than other secondary lymphoid organs[10]. We used intravital microscopy (IVM)[11–13] to visualize the localization of T_{reg} cells and transplanted HSPCs within the bone marrow of live animals.

Prolonged survival of allografts or xenografts without immunosuppression was a criterion used previously to demonstrate immune privilege in sites such as the testis, the eye and the brain[4]. To test if the HSPC niche is an immune-privileged site that meets the above criterion, we first examined if allo-HSPCs transplanted into non-irradiated recipients would survive without any immunosuppression. C57BL/6 (hereafter called B6) C-kit$^+$Sca1$^+$Lin$^-$ (KSL) HSPCs labelled by a membrane dye, 1,1′-dioctadecyl-3,3,3′,3′-tetramethylindodicarbocyanine perchlorate (DiD), were injected into non-irradiated allogeneic BALB/c mice or syngeneic B6 mice. Using IVM, we imaged identical volumes of the recipient skull (calvarial) bone marrow[12,13] on day 30 (Supplementary Fig. 1). Surprisingly, the numbers of surviving KSL cells in the skull bone marrow were comparable in the syngeneic and allogeneic recipients (Fig. 1a, b and Supplementary Fig. 1). Approximately 90% of the donor cells in both recipients were found near the endosteal surface, a location that has been identified as an HSPC niche[6,7] (Fig. 1c). Most KSL cells in both syngeneic and allogeneic groups retained high levels of DiD, indicating that HSPCs did not undergo extensive proliferation. B6-derived cells were not detected in the blood by flow cytometry (data not shown).

Phenotypic analysis showed that surviving donor cells contained a significant fraction of B6 KSL HSPCs and CD150$^+$CD48$^-$KSL haematopoietic stem cells (HSCs) (Supplementary Fig. 2). To determine if they contain functional HSCs, we performed secondary transplant of BALB/c recipient bone marrow into sub-lethally irradiated B6 mice and confirmed long-term multi-lineage reconstitution derived from first donor cells (Supplementary Fig. 3 and Supplementary Table). These results indicate that HSCs with long-term repopulating potential persisted in non-irradiated allogeneic recipients without immunosuppression.

Reciprocal transplant of BALB/c CD150$^+$CD48$^-$Lin$^-$ HSPCs into B6 mice shows that allo-HSPCs can survive in non-irradiated mice regardless of the combination of strains used as donors and recipients (Supplementary Fig. 4).

To explore the possibility that immune ignorance may be the cause of the prolonged survival of allo-HSPCs in non-irradiated recipients, we first examined major histocompatibility complex class I antigen expression and found higher levels of H2-Kb on B6 KSL cells than on spleen CD19$^+$ B cells (data not shown)[14]. To address the possibility that transplantation of KSL cells alone without differentiated cells may fail to evoke a potent allogeneic immune response, we transplanted DiD-labelled KSL cells (5×10^4 per mouse) alone or with non-labelled B6 whole bone marrow cells (5×10^7 per mouse) into non-irradiated BALB/c recipients and detected comparable numbers of surviving DiD-positive B6 cells with or without co-transplantation of whole bone marrow cells (Fig. 1b). The inability of the host to reject these cells is therefore probably not due to immune ignorance, but may reflect active immune regulation.

We next tested if prolonged survival of allogeneic cells in non-irradiated recipients is unique to HSPCs. We simultaneously injected B6 KSL HSPCs and B6 Lin$^+$ differentiated cells labelled with two different dyes, DiD and DiI (1,1′-dioctadecyl-3,3,3′,3′-tetramethylindocarbocyanine perchlorate), into non-irradiated BALB/c or B6 mice. *In vivo* imaging of the skull bone marrows of the allogeneic BALB/c recipients 2 days later showed the presence of both DiI-labelled B6 Lin$^+$ cells and DiD-labelled B6 KSL cells (Fig. 1d), but in contrast

[1]Advanced Microscopy Program, Center for Systems Biology and Wellman Center for Photomedicine, Massachusetts General Hospital, Harvard Medical School, CPZN 8238, 185 Cambridge Street, Boston, Massachusetts 02114, USA. [2]Harvard Stem Cell Institute, 42 Church Street, Cambridge, Massachusetts 02138, USA. [3]Harvard-MIT Division of Health Sciences and Technology, 77 Massachusetts Avenue, E25-519, Cambridge, Massachusetts 02139, USA. [4]Center for Regenerative Medicine, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA. [5]Transplantation Institute and Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. [6]Bone Marrow Transplantation Section, Transplantation Biology Research Center, Massachusetts General Hospital, MGH-East, Bldg.149-5102 13th Street, Boston, Massachusetts 02129, USA. [7]Faculty of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [8]Département de Physique, Génie Physique et Optique & Centre de Recherche Université Laval Robert-Giffard, Québec City, Québec G1J 2G3, Canada. †Present addresses: Department of Regenerative Medicine, Clinical Research Center, National Hospital Organization Nagoya Medical Center, 4-1-1 Sannomaru, Naka-ku, Nagoya Aichi 460-0001, Japan (T.I.S.); Columbia Center for Translational Immunology, Columbia University Medical Center, 650 W. 168th Street, BB 1512, New York, New York 10032, USA (M.S.).

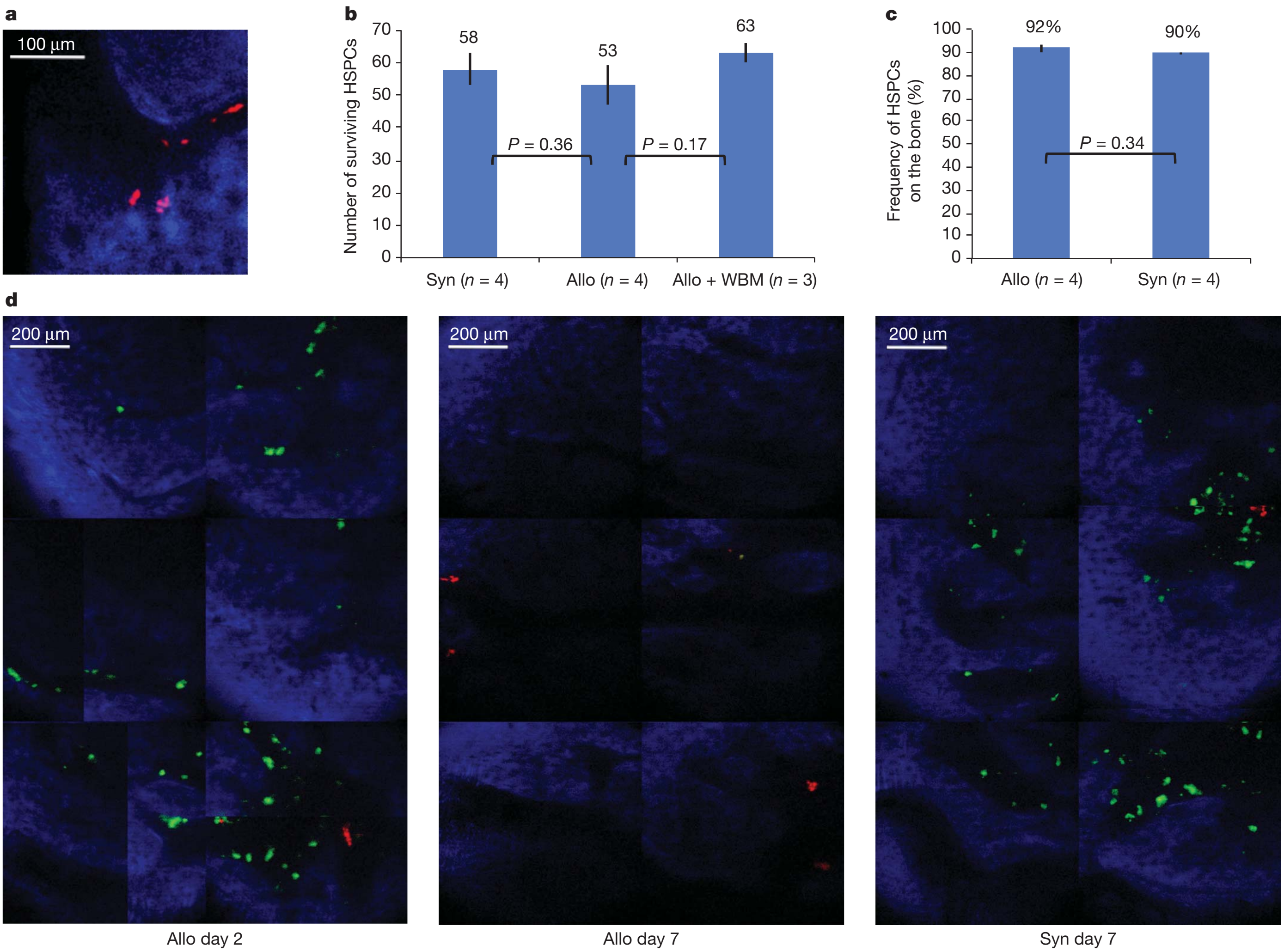


Figure 1 | Allo-HSPCs survive without immunosuppression for 30 days. **a**, *In vivo* imaging of DiD-labelled B6 KSL HSPCs (red) surviving on the endosteal surface of the skull bone marrow of BALB/c recipients for 30 days without immunosuppression. Blue, bone. Black area is bone marrow cavity. **b**, Comparable numbers of DiD-labelled B6 KSL HSPCs (5×10^4 cells per mouse) in the skull bone marrow on day 30 after transplantation into B6 (Syn) mice or BALB/c (Allo) mice ($n = 4$ recipients). Co-transplantation of non-labelled B6 whole bone marrow (WBM) cells did not affect the number of surviving HSPCs ($n = 3$ recipients). Cell numbers were counted in identical volumes of the skull covering 1,650 μm in *x*, 2,310 μm in *y*, and 150 μm in *z*. Error bars indicate standard error. **c**, Most B6 donor KSL cells localize within 15 μm of the endosteal surface in both B6 and BALB/c recipients on day 30 after transplantation. Error bars indicate standard error ($n = 4$). **d**, Montage pictures of the skull bone marrows of non-irradiated BALB/c mice or B6 mice 7 days after transplantation of DiD-labelled B6 KSL HSPCs (red) and DiI-labelled B6 Lin^+ differentiated cells (green). Blue, bone. Black area is bone marrow cavity. Image size was 880 μm (*x*) × 1,320 μm (*y*) around the coronal sutures and central vein.

to the KSL cells, the Lin^+ cells were distributed farther away from the endosteal surface (Supplementary Fig. 5). Notably, on day 7, Lin^+ cells were virtually undetectable, whereas the HSPCs persisted (Fig. 1d). In the syngeneic B6 recipients, both populations survived for 7 days (Fig. 1d).

Prolonged survival of allo-HSPCs in non-irradiated recipients is consistent with earlier observations of prolonged survival of transplanted allogeneic/xenogeneic grafts in other immune-privileged sites[4], supporting our hypothesis that the HSPC niche also has immune-privileged mechanisms that shield the transplanted allo-HSPCs, but not their differentiated counterpart, from allogeneic immune response.

Next, we examined the relevance of T_{reg} cells to immune privilege in the HSPC niche. We confirmed that the frequency of FoxP3 T_{reg} cells in CD4 T cells is higher in the bone marrow (23%) than in the spleen (13%) and lymph nodes (13%) (data not shown). As the trafficking of T_{reg} cells to the bone marrow critically depends on the stem-cell chemo-attractant SDF-1 (ref. 10), which is expressed at high levels by cellular components of the HSPC niche[13,15–17], we proposed that T_{reg} cells will accumulate in the HSPC niche and render it immune suppressive. *In vivo* three-dimensional imaging of the intact skull and mechanically thinned proximal tibia of FoxP3–GFP mice, whose $FoxP3^+$ cells specifically express the green fluorescent protein (GFP), showed that 58% (203 of 350) and 72% (137 of 191), respectively, of T_{reg} cells were within 15 μm of the endosteal surface (Fig. 2a, b). Preferential localization of T_{reg} cells on the endosteal surface was also confirmed by analysis of two-dimensional histological sections (Fig. 2b and Supplementary Fig. 6).

Next, we examined the relative spatial distribution of T_{reg} cells and transplanted HSPCs by *in vivo* three-dimensional imaging. Notably, >80% of the DiD-labelled B6 KSL cells homed within 20 μm of T_{reg} cells on the endosteal surface 24 h after intravenous injection into FoxP3–GFP B6 mice (Fig. 2a, c). In contrast, Lin^+ differentiated cells were much less likely to home close to T_{reg} cells (Fig. 2c). Close proximity of transplanted KSL cells and T_{reg} cells on the endosteal surface was confirmed by histological analysis of the tibia and femur trabecular bone marrow (Supplementary Fig. 6). We also visualized the quiescent HSCs after administering the chemotherapy drug 5-fluorouracil (5-FU), which selectively killed dividing cells. *In vivo* imaging of the skull bone marrow 48 h after 5-FU treatment showed that 82% (36 of 44) of the 5-FU-resistant DiD-labelled KSL cells were within 20 μm of

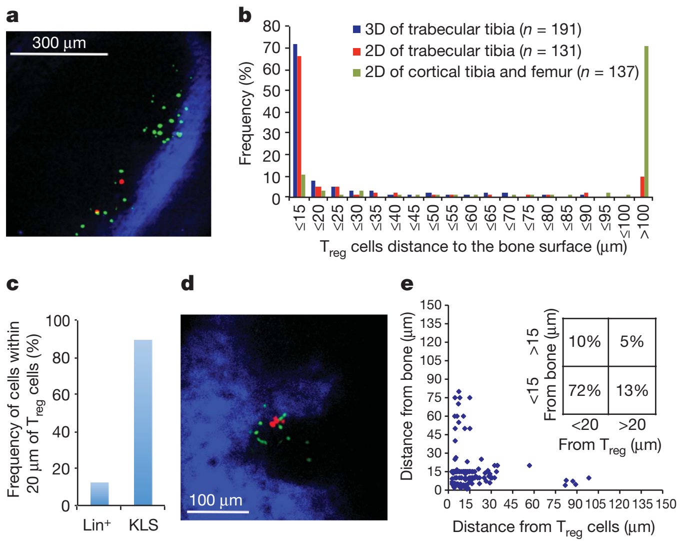


Figure 2 | FoxP3 T_{reg} cells accumulate on the endosteal surface, form clusters around adoptively transferred HSPCs and around allo-HSPCs that survive after 30 days. **a**, *In vivo* imaging of the tibia trabecular bone marrow of B6 FoxP3–GFP mice transplanted with DiD-labelled B6 KSL HSPCs. FoxP3–GFP T_{reg} cells (green) accumulate on the endosteal surface. Of DiD-labelled B6 KSL HSPCs (red) 81% (83 of 103) home within 20 μm of T_{reg} cells. Blue, bone. Black area is bone marrow cavity. The maximal imaging depth of the tibia trabecular bone marrow was approximately 150 μm below the bone surface, whereas the size of the tibia bone marrow space confined by anastomosing trabeculae ranged between 50 and 250 μm based on the micro-computed tomography analysis (data not shown). **b**, Histogram depicting the distance of FoxP3–GFP T_{reg} cells to the endosteal surface. Three-dimensional (3D) results were obtained by IVM, whereas two-dimensional (2D) results were obtained from histological sections. **c**, Frequency of adoptively transferred KSL cells (89% (338 of 380)) and Lin^+ cells (12% (42 of 330)) homing within 20 μm of FoxP3–GFP T_{reg} cells in the skull bone marrow. **d**, *In vivo* imaging of 5-FU-resistant DiD-labelled KSL HSPCs (red) forming clusters with FoxP3 T_{reg} cells (green) on the endosteal surface. Blue, bone. Black area is bone marrow cavity. **e**, Spatial distribution of B6 KSL HSPCs surviving in the skull bone marrow of non-irradiated BALB/c FoxP3–GFP mice for 30 days relative to the endosteal surface and T_{reg} cells. Measurements were taken from three-dimensional *in vivo* skull bone marrow image stacks.

the FoxP3–GFP T_{reg} cells (Fig. 2d), forming clusters with T_{reg} cells on the endosteal surface. Similarly, T_{reg} cells accumulated around allo-HSPCs that survived on the endosteal surface of the skull bone marrow and trabecular bone marrow of long bones in non-irradiated mice for 30 days without immunosuppression (Fig. 2e and Supplementary Fig. 7). In cortical bone marrow of long bones, only 32% (38 of 120) of allo-HSPCs and 11% (15 of 137) of T_{reg} cells were within 15 μm of the endosteal surface, but proximity of T_{reg} cells to allo-HSPCs was still noted irrespective of their distance to the endosteal surface (Fig. 2b and Supplementary Fig. 7).

To examine whether T_{reg} cells have a critical role in providing immune privilege to the HSPC niche, we examined if depletion of T_{reg} cells resulted in the rejection of allo-HSPCs without immunosuppression. Using FoxP3–GFP diphtheria toxin receptor (DTR) mice, the T_{reg} cells of which express DTR and can be specifically depleted by diphtheria toxin treatment (data not shown), we observed a 70% reduction in the number of surviving donor cells in diphtheria-toxin-treated versus control-vehicle-treated recipients (Fig. 3a). Depletion of T_{reg} cells by anti-CD25 antibody also led to a 90% reduction in the number of surviving donor cells compared with control-antibody-treated

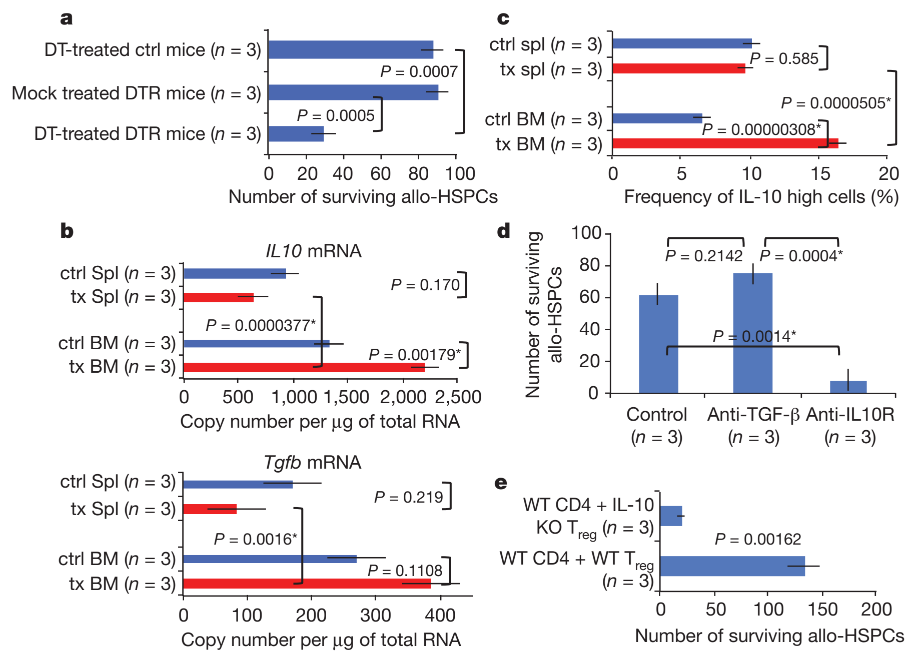


Figure 3 | Bone marrow T_{reg} cells are critical in suppressing the rejection of allo-HSPCs in immune-competent recipients in an IL-10-dependent manner. **a**, The number of DiD-labelled BALB/c $CD150^+CD48^-Lin^-$ HSPCs in the skull bone marrow of diphtheria-toxin (DT)-treated FoxP3–GFP DTR mice (B6N9), mock-treated FoxP3–GFP DTR mice (B6N9), and diphtheria-toxin-treated control mice (B6N9). Cells were counted in identical volumes 2,640 μm (x) × 3,330 μm (y) × 150 μm (z) in size. Error bars indicate standard error ($n = 3$). **b**, qRT–PCR analysis of $FoxP3^+$ T_{reg} cells isolated from B6 FoxP3–GFP recipients 6 days after transplantation of BALB/c $CD150^+CD48^-Lin^-$ HSPCs without immunosuppression. Results are expressed as copy number per μg of total RNA. Error bars indicate standard error ($n = 3$). BM, bone marrow. **c**, Intracellular staining of IL-10 in FoxP3–GFP T_{reg} cells in non-irradiated mice with and without allo-HSPC transplants. Error bars indicate standard error ($n = 3$). **d**, Quantification of DiD-labelled B6 KSL cells in the skull bone marrow of anti-IL-10R, anti-TGF-β or control IgG antibody-treated BALB/c mice after transplantation of DiD-labelled B6 KSL cells without immunosuppression. Error bars indicate standard error ($n = 3$). Cells were counted in identical volumes 2,620 μm (x) × 2,200 μm (y) × 150 μm (z) in size. **e**, Quantification of DiD-labelled BALB/c $CD150^+CD48^-Lin^-$ HSPCs in the skull bone marrow of B6 RAG2 knockout mice reconstituted with wild-type (WT) $CD4^+CD25^-$ T cells and IL-10 knockout $CD4^+CD25^+$ T_{reg} cells or with wild-type $CD4^+CD25^-$ T cells and wild-type $CD4^+CD25^+$ T_{reg} cells. Cells were counted in identical volumes 2,640 μm (x) × 3,330 μm (y) × 150 μm (z) in size. Error bars indicate standard error ($n = 3$).

recipients (Supplementary Fig. 8). Quantitative PCR with reverse transcription (qRT–PCR) analysis of $CD4^+FoxP3^-$ T cells and $CD8^+$ T cells isolated from the bone marrow showed increased levels of TNF-α and IL-4 expression in diphtheria-toxin-treated versus control B6N9 FoxP3–GFP DTR mice transplanted with BALB/c HSPCs (Supplementary Fig. 9). Together these results indicate that T_{reg} cells critically provide the HSPC niche with immune privilege and that T_{reg}-cell-mediated suppression of bone marrow CD4 and CD8 T-cell activation is at least in part responsible for the survival of allo-HSPCs in non-irradiated hosts without immunosuppression.

We also performed qRT–PCR analysis of T_{reg} cells from B6 FoxP3–GFP recipients 6 days after transplantation of BALB/c $CD150^+CD48^-Lin^-$ HSPCs without immunosuppression. *Tgfb* mRNA level in bone marrow FoxP3 T_{reg} cells was fourfold higher than in spleen T_{reg} cells of transplanted recipients, but this level was comparable to that in bone marrow T_{reg} cells of non-transplanted recipients (Fig. 3b). In contrast, *IL10* mRNA level in bone marrow T_{reg} cells of transplanted recipients was significantly higher compared with control bone marrow T_{reg} cells without transplant, and was also significantly higher compared with spleen T_{reg} cells of transplanted mice (Fig. 3b). Increased expression of IL-10 at the protein level in bone marrow FoxP3 T_{reg} cells was also confirmed using fluorescence-activated cell sorting (FACS) analysis after intracellular staining (Fig. 3c). Consistent with these findings, anti-IL-10R antibody treatment led to ~90% reduction of the number of surviving allo-HSPCs compared with control IgG antibody treatment, whereas anti-TGF-β antibody treatment did not have a significant effect on the number of donor cells (Fig. 3d).

To test the specific role of IL-10 in T_{reg} cells, we used as transplant recipients non-irradiated B6 RAG2 knockout mice reconstituted 1 day earlier with B6 IL-10 knockout $CD4^+CD25^+$ T_{reg} cells or B6 wild-type $CD4^+CD25^+$ T_{reg} cells together with B6 wild-type $CD4^+CD25^-$ T cells. The numbers of DiD-labelled BALB/c HSPCs surviving for 7 days in the recipients with IL-10 knockout T_{reg} cells were markedly lower than that in control recipients with wild-type $CD4^+CD25^+$ T cells (Fig. 3e). Together, the data indicate that IL-10 from T_{reg} cells is critical for T_{reg}-mediated immune protection of allo-HSPCs in non-irradiated recipients.

We show here that in anatomical locations where immune activity is otherwise known to occur[8,9], discrete sites of immune privilege may exist and contribute to a somatic stem-cell niche. Prolonged survival of allo-HSPCs without immunosuppression is surprising and seems contrary to the clinical experience that indicates strong immunosuppression is required to prevent rejection in allogeneic bone marrow transplantation[18]. We propose that our data may not be in conflict with that well-defined experience because we are observing events at the level of the HSPC in non-irradiated hosts, whereas the clinical transplant measure of engraftment is mature cell chimaerism. In transplantation without immunosuppression, mature cells are susceptible to immune attack and low levels of HSPC chimaerism in the relative absence of proliferation may go undetected by conventional methods.

T_{reg} cells accumulate in the HSPC niche and may provide the HSPC niche with immune privilege mechanisms, enabling transplanted allo-HSPCs to escape from allogeneic rejection. Immune privilege mechanisms of the HSPC niche will shield endogenous HSPCs from autoimmunity or excessive inflammation, and will even help malignant cells to evade host immunity. This work raises the possibility of niches in other tissues serving as immune-privileged sites.

METHODS SUMMARY

All mice were housed according to IACUC guidelines and used for experiments when 8–14-weeks old. Mice were anaesthetized and prepared for *in vivo* imaging as described previously[12,13]. The mouse was held in a heated tube mounted on a precision 3-axis motorized stage (Sutter MP385). All mice were imaged with a custom-built confocal two-photon hybrid microscope specifically designed for live-animal imaging (see Methods). Using the crossing of the central vein and coronal sutures as landmarks, we imaged identical areas of the skull (~1,650 μm × 2,310 μm) encompassing most of the parasagittal bone marrow cavities[12,13]. We acquired three-dimensional stacks consisting of 31 optical sections with 5 μm *z* spacing, which provide image volumes 150 μm in depth, reaching approximately 40–60% into the bone marrow cavity[12].

Image processing and HSPC-microenvironment distance measurements were obtained using ImageJ[12]. A two-tailed *t*-test was applied in the two-group analysis. P values ≤ 0.05 were considered to be statistically significant. In the three and four group comparisons, we performed one-way analysis of variance followed by an analysis of the three pairwise differences of interest using Bonferroni correction so that P values $\leq 0.05/3$ were considered statistically significant.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 September 2009; accepted 27 April 2011.

1. Fuchs, E., Tumbar, T. & Guasch, G. Socializing with the neighbors: stem cells and their niche. *Cell* **116,** 769–778 (2004).
2. Li, L. & Xie, T. Stem cell niche: structure and function. *Annu. Rev. Cell Dev. Biol.* **21,** 605–631 (2005).
3. Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118,** 149–161 (2004).
4. Niederkorn, J. Y. See no evil, hear no evil, do no evil: the lessons of immune privilege. *Nature Immunol.* **7,** 354–359 (2006).
5. Adams, G. B. & Scadden, D. T. The hematopoietic stem cell in its place. *Nature Immunol.* **7,** 333–337 (2006).
6. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425,** 836–841 (2003).
7. Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425,** 841–846 (2003).
8. Feuerer, M. *et al.* Bone marrow as a priming site for T-cell responses to blood-borne antigen. *Nature Med.* **9,** 1151–1157 (2003).
9. Joffre, O. *et al.* Prevention of acute and chronic allograft rejection with $CD4^+CD25^+Foxp3^+$ regulatory T lymphocytes. *Nature Med.* **14,** 88–92 (2008).
10. Zou, L. *et al.* Bone marrow is a reservoir for $CD4^+CD25^+$ regulatory T cells that traffic through CXCL12/CXCR4 signals. *Cancer Res.* **64,** 8451–8455 (2004).
11. Mazo, I. B. *et al.* Hematopoietic progenitor cell rolling in bone marrow microvessels: parallel contributions by endothelial selectins and vascular cell adhesion molecule 1. *J. Exp. Med.* **188,** 465–474 (1998).
12. Lo Celso, C. *et al.* Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche. *Nature* **457,** 92–96 (2009).
13. Sipkins, D. A. *et al.* *In vivo* imaging of specialized bone marrow endothelial microdomains for tumour engraftment. *Nature* **435,** 969–973 (2005).
14. Huang, Y. *et al.* Matching at the MHC class I K locus is essential for long-term engraftment of purified hematopoietic stem cells: a role for host NK cells in regulating HSC engraftment. *Blood* **104,** 873–880 (2004).
15. Jung, Y. *et al.* Regulation of SDF-1 (CXCL12) production by osteoblasts; a possible mechanism for stem cell homing. *Bone* **38,** 497–508 (2006).
16. Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25,** 977–988 (2006).
17. Méndez-Ferrer, S. *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466,** 829–834 (2010).
18. Welniak, L. A., Blazar, B. R. & Murphy, W. J. Immunobiology of allogeneic hematopoietic stem cell transplantation. *Annu. Rev. Immunol.* **25,** 139–170 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by Bullock fellowship (to J.F.), the Harvard Stem Cell Institute and DoD W81XWH-10-1-0217 (to J.F. and C.P.L.), NIH HL097748 (to C.P.L.), HL97794 (to D.T.S.), CA111519 (to M.S.), AI041521 (to T.B.S), EMBO (to C.L.C.), HFSP (to C.L.C.), philanthropic sources (to D.T.S. and C.L.C.), and the National Institutes of Health. We acknowledge J. Zhao and D. Cao for cell sorting.

Author Contributions J.F. conceived the overall project, designed and performed experiments, analysed data and wrote the manuscript. J.W. and A.L.C. contributed to intravital imaging. L.S. and J.F. performed histology analysis. P.P. and J.F. performed qRT–PCR analysis. R.L. and J.F. performed FACS analysis using intracellular staining. W.G. performed mouse breeding. J.F., H.T. and T.S. performed HSPC isolation. D.C. generated software for IVM. T.I.S., C.L.C., M.S., T.B.S. and D.T.S. provided advice on the design of experiments. D.T.S. edited the manuscript. C.P.L. helped design experiments, supervised the overall study and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.F. (jfujisaki@partners.org) or C.P.L. (lin@helix.mgh.harvard.edu).

METHODS

Mouse. FoxP3–GFP knock-in mice (C57BL/6 or BALB/c background) were provided by T. Strom's laboratory. FoxP3–GFP diphtheria toxin receptor (DTR) mice were provided by Rudensky's laboratory. C57BL/6 mice, BALB/c mice and IL-10 knockout mice were purchased from Jackson laboratory. RAG2 knockout mice were purchased from Taconic. All animal experiments were performed in compliance with institutional guidelines and approved by the Subcommittee on Research Animal Care (SRAC) at Massachusetts General Hospital.

Haematopoietic stem-cell isolation. Whole bone marrow was obtained from C57BL/6 mice or BALB/c mice by crushing the femur, tibia, iliac, humerus and vertebral bones. Cells were pooled and washed with PBS, then incubated in a lineage cocktail consisting of biotinylated B220, Mac1, GR-1, CD3a, CD8a, CD4 and Ter119 antibodies. After washing, cells were incubated with MACS-SA beads (Miltenyi) following the manufacturer's protocol, washed again, then separated on an LD depletion column in a MidiMACS separation unit to remove lineage-specific cell populations. Lineage-negative cells from B6 mice were then stained with C-kit and Sca1 antibodies conjugated with allophycocyanin (APC) and phycoerythrin (PE), respectively. Lineage-negative cells from BALB/c mice were stained with CD48-APC and CD150-PE antibodies. Subsequently, C-kit$^+$Sca1$^+$Lin$^-$ cells or CD150$^+$CD48$^-$Lin$^-$ cells were isolated by FACS sorting machine (BD FACSAria). Cells were then stained with DiD and injected into recipient mice for intravital confocal microscopy.

Fluorescence cell labelling. C-kit$^+$Sca1$^+$Lin$^-$ cells, Lin$^+$ cells, and CD150$^+$ CD48$^-$Lin$^-$ cells were fluorescently labelled by incubation with the dialkylcarbocyanine membrane dyes, 'DiD' or 'DiI'. Cells in culture media were incubated with 10 μM dye for 30 min at 37 °C. Cells were then centrifuged and the pellet washed in PBS three times before injection in animals.

***In vivo* imaging.** *In vivo* imaging of the bone marrow through the intact skull bone of live mice was first demonstrated elsewhere[11]. Here we used a video-rate laser scanning hybrid confocal/two-photon microscope that is specifically designed and optimized for live animal imaging[12,13]. A polygon-based scanning engine allows simultaneous multi-channel image acquisition at video rate (30 frames per second), a feature that is particularly useful for imaging moving objects. Fast scanning speed, together with a precision computer-controlled *xyz* stage, also facilitate surveying large tissue volumes in three dimensions and searching for rare cells in the bone marrow.

The mice were anaesthetized with intraperitoneal injection of ketamine/xylazine, and a small incision was made in the scalp to expose the underlying skull. The skull bone was kept intact. Second harmonic microscopy was used to visualize the bone and to identify the major anatomical landmarks such as the central vein and the coronal suture. Bone is rich in type-1 collagen, which was imaged by second harmonic generation using 840 nm excitation and 420 nm detection. Using the crossing of the central vein and coronal sutures as landmarks, we imaged identical areas of the skull (~1,650 μm × 2,310 μm) encompassing most of the parasagittal bone marrow cavities[12,13]. The maximum imaging depth was approximately 150 μm below the outer bone surface, allowing us to penetrate between 40% and 60% of bone marrow cavity space in over 75% of our measurements[12]. DiD signal was excited with a 638-nm helium–neon laser (Radius, Coherent Inc.) and detected with a photomultiplier tube through a 695 ± 27.5-nm band-pass filter (Omega Optical). GFP signal was excited with a 491-nm solid state laser (Dual Calypso, Cobolt AB) and detected with a photomultiplier tube through a 528 ± 19-nm band-pass filter (Semrock). Second harmonic generation signals from the bone were excited with an 840-nm titanium–sapphire laser (Spectra Physics) and detected with a photomultiplier tube through a 435 ± 20-nm band-pass filter (Semrock). Each image consisted of 30 frames averaged together. Imaging duration per mouse was approximately 2 h.

For imaging the tibia bone marrow, the outer skin was removed from the leg and tendons around the joints severed to detach adjoining muscles. The saphenous and medial marginal vessels were kept intact. Upon exposure of the tibia, the medial bone surface was then gently polished with 400 grit silicon carbide wet grinding paper (Struers) while being irrigated with PBS and viewed with a surgical microscope until the fine structures of the trabecular bone became visible under the thinned cortical bone. The area was cleaned with a PBS flush and imaged using the same set-up as described above for imaging the skull bone marrow. Maximal imaging depth is approximately 150 μm below the bone surface.

Histological analysis. Femurs and tibias were fixed in 4% paraformaldehyde for 4 h at room temperature, cryoprotected in 30% sucrose for 2–3 h, immersed in OCT compound (Sakura) and frozen in liquid nitrogen in the horizontal orientation. Thick frozen sections of undecalcified bone were obtained, whereby a layer of compact bone was removed from the top and the bottom of the specimen, leaving both sides of the bone marrow cavity accessible for imaging. The specimens were subsequently thawed, washed in PBS, embedded in agarose and imaged using the same set-up as described above.

Image processing and analysis. Image processing and HSPC-microenvironment distance measurements were obtained using ImageJ as described previously[12]. A two-tailed *t*-test was applied in the two-group analysis. P values ≤ 0.05 were considered statistically significant. In the three and four group comparison, we performed one-way analysis of variance followed by an analysis of the three pairwise differences of interest using Bonferroni correction so that P values $\leq 0.05/3$ were considered statistically significant.

Flow cytometry analysis of transplanted donor HSPCs. Bone marrow cells were isolated from non-irradiated BALB/c mice 30 days after transplantation with B6 KSL cells (1×10^5 cells per mouse), depleted of Lin$^+$ cells using EasySep (StemCell Technologies). Cells were subsequently stained with PE–H2Kb, APC–CD48, APC–Cy7–C-kit, Pacific Blue–Sca1, Pacific Orange–Streptavidin, and PE–Cy7–CD150 antibodies. Cells were analysed by FACS (BD FACSAria) with FlowJo software.

Secondary transplantation. Bone marrow cells were isolated from non-irradiated BALB/c mice (CD45.2) 30 days after transplantation with B6 KSL cells (CD45.2) (5×10^4 cells per mouse), depleted of CD4 and CD8 T cells, and intravenously transplanted into 7.0-Gy or 8.4-Gy-irradiated B6 SJL mice (CD45.1). Flow cytometry analysis of peripheral blood cells was performed 30 days or 240 days after secondary transplantation, following the staining of cells by PE-Cy5.5-conjugated CD45.2 antibody, PE-conjugated H2Kb, and APC-conjugated CD3, B220, CD11b, or Gr1 antibodies (all from BD Bioscience). Cells were analysed by FACS machine (BD FACSAria). The data were analysed by FlowJo.

Flow cytometry analysis of FoxP3–GFP T_{reg} cells. Whole bone marrow cells, spleen cells and lymph node cells were isolated from FoxP3–GFP mice and stained with anti-CD4, CD25, NK1.1 antibodies (eBioscience). Intracellular staining for IL-10 was performed according to the manufacturer's protocol (eBioscience).

5-FU treatment. Eight-week-old mice were injected with 5×10^4 million KSL HSPCs per mouse labelled by DiD. The skull bone marrow was imaged 24 h later by IVM. At the completion of the initial imaging session, the skull was flushed with sterile saline, and the scalp skin flaps were sewn closed using 5-0 non-absorbable nylon sutures. Mice were injected with 5-FU (250 mg per kg body weight, Sigma) intravenously[3]. Sutures were removed 48 h after 5-FU treatment[3], and the area flushed and swabbed with sterile saline for subsequent skull bone marrow imaging.

Diphtheria toxin treatment. Non-irradiated B6N9 FoxP3–GFP DTR mice were transplanted with DiD-labelled BALB/c CD150$^+$CD48$^-$Lin$^-$ HSPCs (4×10^4 cells per mouse). Diphtheria toxin treatment (50 μg per kg of mouse) or mock treatment was given intraperitoneally every other day from day 0. *In vivo* imaging of the skull bone marrows of the recipients were taken on day 11.

***In vivo* antibody treatment.** Non-irradiated BALB/c mice were transplanted with DiD-labelled B6 KSL HSPCs. On day 0 and day 2, rat anti-CD25 antibody (1 mg per mouse, PC-61, Bio X Cell) or control rat IgG antibody (1 mg per mouse, Bio X Cell) was given intravenously. Mouse anti-IL-10R antibody, anti-TGF-β antibody or control IgG antibody treatment (200 μg per mouse, Bio X Cell) was given intraperitoneally every other day from day −1 until day 7.

RNA extraction and cDNA synthesis. Isolated cells were homogenized using a 25-gauge syringe needle. Total RNA was extracted using Invitrogen's Purelink micro to midi kit and the RNA concentration was measured on a NanoDrop ND 1000 spectrophotometer (Nanodrop). Total RNA (100 ng) was converted into cDNA using a TaqMan Reverse Transcription kit (Applied Biosystems)[19].

Quantitative TaqMan PCR with reverse transcription. The ABI PRISM 7900HT Sequence Detection System was used for qRT–PCR analysis. All primer–probe sets were custom designed. Custom-designed primer–probe sets were validated by serially diluting cDNA isolated from cells expressing the target and verifying the slope, and by sequencing the amplicon. Taqman Fast PCR Master Mix was purchased from Qiagen. Amplification was carried out in a total volume of 25 μl for 40 cycles of 3 s at 95 °C, 30 s at 60 °C. Initial denaturation was performed for 3 min at 95 °C. Target gene expression was normalized by total RNA quantity used for cDNA synthesis.

Immune reconstitution in RAG2 knockout mice. CD4$^+$CD25$^-$ cells and CD4$^+$CD25$^+$ cells were isolated from B6 mice and B6 IL-10 knockout mice. B6 RAG2 knockout mice were adoptively transferred with B6 wild-type CD4$^+$CD25$^-$ T cells (1.5 million per mouse) and B6 IL-10 knockout CD4$^+$CD25$^+$ T_{reg} cells (0.5 million per mouse), or wild-type CD4$^+$CD25$^-$ T cells (1.5 million per mouse) and B6 CD4$^+$CD25$^+$ T_{reg} cells (0.5 million per mouse). One day later, DiD-labelled BALB/c CD150$^+$CD48$^-$Lin$^-$ HSPCs were transplanted.

19. Bustin, S. A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* **25**, 169–193 (2000).

doi:10.1038/nature10107

A somitic Wnt16/Notch pathway specifies haematopoietic stem cells

Wilson K. Clements[1], Albert D. Kim[1], Karen G. Ong[1], John C. Moore[2], Nathan D. Lawson[2] & David Traver[1]

Haematopoietic stem cells (HSCs) are a self-renewing population of cells that continuously replenish all blood and immune cells during the lifetime of an individual[1,2]. HSCs are used clinically to treat a wide array of diseases, including acute leukaemias and congenital blood disorders, but obtaining suitable numbers of cells and finding immune-compatible donors remain serious problems. These difficulties have led to an interest in the conversion of embryonic stem cells or induced pluripotent stem cells into HSCs, which is not possible using current methodologies. To accomplish this goal, it is critical to understand the native mechanisms involved in the specification of HSCs during embryonic development. Here we demonstrate in zebrafish that Wnt16 controls a novel genetic regulatory network required for HSC specification. Non-canonical signalling by Wnt16 is required for somitic expression of the Notch ligands *deltaC* (*dlc*) and *deltaD* (*dld*), and these ligands are, in turn, required for the establishment of definitive haematopoiesis. Notch signalling downstream of Dlc and Dld is earlier than, and distinct from, known cell-autonomous requirements for Notch, strongly suggesting that novel Notch-dependent relay signal(s) induce the first HSCs in parallel to other established pathways. Our results demonstrate that somite-specific gene expression is required for the production of haemogenic endothelium.

We wished to define better the role, if any, of Wnt signalling in HSC specification during embryonic development. Although Wnt signalling can exert strong effects on adult HSCs, an *in vivo* regulatory function is controversial, and the role of Wnt signalling in establishing haematopoiesis during development is unclear[2]. Wnt signalling pathways have been loosely grouped into two families: canonical and non-canonical[2,3]. Canonical signalling is defined by stabilization of β-catenin, which in cooperation with lymphoid-enhancer-binding factor/T-cell factor (Lef/Tcf) DNA-binding proteins, activates transcription of Wnt target genes[2,3]. Non-canonical signalling is β-catenin/Tcf-independent, with less well-characterized intracellular pathways[3]. Disruption of canonical signalling in zebrafish by targeted deletion of *wnt3a* results in HSC deficits[2], but these animals have massive morphological defects, including near absence of the caudal tissues[4] where HSCs arise during embryogenesis. Canonical Wnt signalling by unknown ligands also seems to have a role in maintaining and/or expanding very early HSCs in cooperation with prostaglandins[5]. At present, no studies have demonstrated an absolute requirement for Wnt signalling in the earliest specification of recognizable HSCs, and no requirement for β-catenin/Tcf-independent, non-canonical signalling has been reported.

We searched for candidate Wnt proteins expressed near pre-haematopoietic mesoderm and identified a previously uncharacterized zebrafish *wnt16* orthologue (Supplementary Fig. 1) expressed in the dorso-anterior portion of more rostral somites from 10 h post fertilization (h.p.f.; tailbud stage) to 24 h.p.f. (Supplementary Fig. 2), the time frame when pre-haematopoietic mesoderm transitions to recognizable HSC precursors in the dorsal aorta. *Wnt16* is conserved across vertebrate phyla (Supplementary Fig. 1), and the human form was originally identified as a gene inappropriately expressed in pre-B-acute lymphoblastic leukaemia (ALL) cells containing the E2A-PBX1 t(1;19) translocation product[6]. In zebrafish, two *wnt16* splice variants are produced, only one of which is active (Supplementary Figs 1a, b and 3 and Supplementary Table 1).

Knock down of Wnt16 by injection of either of two ‘splice-blocking’ morpholinos (W16MO) caused a striking haematopoietic phenotype. Morpholinos alone or in combination reduced levels of functional *wnt16* mRNA (Supplementary Figs 1a and 4), yielding highly similar phenotypes in comparison to a 5-base-pair mismatch control morpholino, which had no effect (Supplementary Figs 4 and 5). Whole-mount *in situ* hybridization (WISH) revealed that HSC precursors, a population of *runx1*$^{+}$ cells in the ventral floor of the dorsal aorta[1,7,8], as in mouse[9], are absent in embryos injected with W16MO at 24 h.p.f. (red arrowhead; Fig. 1a, b and Supplementary Table 2). By 33 h.p.f., HSCs can be observed as a population of *cmyb*$^{+}$ cells between the dorsal aorta and posterior cardinal vein[1], and this population is also absent in W16MO-injected embryos (red arrowhead, Fig. 1c, d and Supplementary Table 2). HSCs observable in living transgenic animals carrying *GFP* under the control of the *cd41* (also known as *itga2b*) promoter[10–12] are absent or reduced in *wnt16* morphants (red arrowheads, Fig. 1e, f; see also Supplementary Movies 1 and 2 and Supplementary Table 2), as are unique double-positive cells in *kdrl:RFP* and *cmyb:GFP* double transgenics[7] (Fig. 1g, h, yellow cells; see also Supplementary Table 2), whereas unrelated GFP-labelled multiciliate cells in the pronephros are unaffected (yellow arrowheads in Fig. 1e, f and green cells in Fig. 1g, h). Generation of T lymphocytes requires an HSC precursor[10,11,13], providing a useful readout for whether HSCs have been specified or not. In *cd41:GFP* transgenic animals, GFP^{+} HSCs retain residual fluorescence as they differentiate into the first thymic lymphocytes around 3 days post fertilization (d.p.f.)[10], and these thymic immigrants are greatly reduced in *wnt16* morphants (Supplementary Movies 3 and 4). Accordingly, *wnt16* morphants lack *rag1*$^{+}$ T lymphocytes at 4.5 d.p.f. (blue arrows, Fig. 1i, j and Supplementary Table 2), despite intact thymic epithelium (Supplementary Fig. 6). Together, our results indicate that Wnt16 is required for the specification of HSCs during embryonic development.

To determine the specificity of W16MO-induced defects, we examined the integrity of non-haematopoietic tissues by morphology and expression of marker genes (Fig. 2 and Supplementary Table 3). *Wnt16* morphants are grossly normal (Fig. 2a, b), with intact and functional vasculature, as evidenced by beating hearts and circulating primitive, non-HSC-derived erythroid cells, although there appear to be subtle patterning defects in the number and position of intersegmental vessels. WISH revealed that W16MO-injected embryos have primitive blood (*gata1*; Fig. 2c, d and Supplementary Table 3), somites (*myod*; Fig. 2e, f and Supplementary Table 3), vasculature (*tll1*, *cdh5* and *kdrl* (also called *flk1*); Fig. 2g–l and Supplementary Table 3), dorsal aorta (*efnb2a*; Fig. 2m, n; *notch1b*, *notch3*, *dll4*; Supplementary Fig. 7a–d, o, p and Supplementary Tables 3 and 5), hypochord (*col2a1a*; Fig. 2o, p and Supplementary Table 3), notochord and floorplate (*shha*; Fig. 2q, r and Supplementary Table 3), and pronephros (*cdh17*; Fig. 2s, t and

[1]Department of Cellular and Molecular Medicine and Section of Cell and Developmental Biology, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0380, USA. [2]Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.

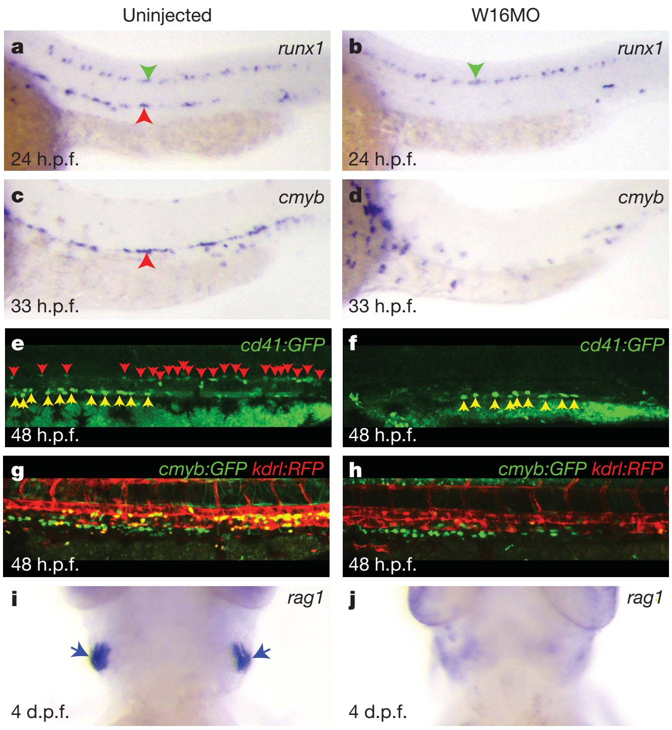


Figure 1 | ***Wnt16* is required for the specification of HSCs.** **a**–**d**, Expression of the HSC markers *runx1* (**a**, **b**) and *cmyb* (**c**, **d**). **e**–**h**, Fluorescently labelled HSCs in *cd41:gfp* (**e**, **f**) and *cmyb:GFP;kdrl:RFP* transgenics (**g**, **h**). **i**, **j**, Expression of the lymphocyte marker *rag1*. Embryos are either uninjected (left column) or injected with 5 ng W16MO (right column). Red arrowheads identify the aorta region (**a**, **c**) or individual HSCs (**e**). Green arrowheads identify unaffected *runx1*$^+$ neurons (**a**, **b**). Yellow arrowheads identify unaffected GFP$^+$ multiciliate cells of the pronephros (**e**, **f**). Yellow cells are HSCs (**g**). Blue arrows identify thymic T cells (**i**). **a**–**h**, Dorsal up, anterior left. **i**–**j**, Ventral views, anterior up. Original magnification at ×200.

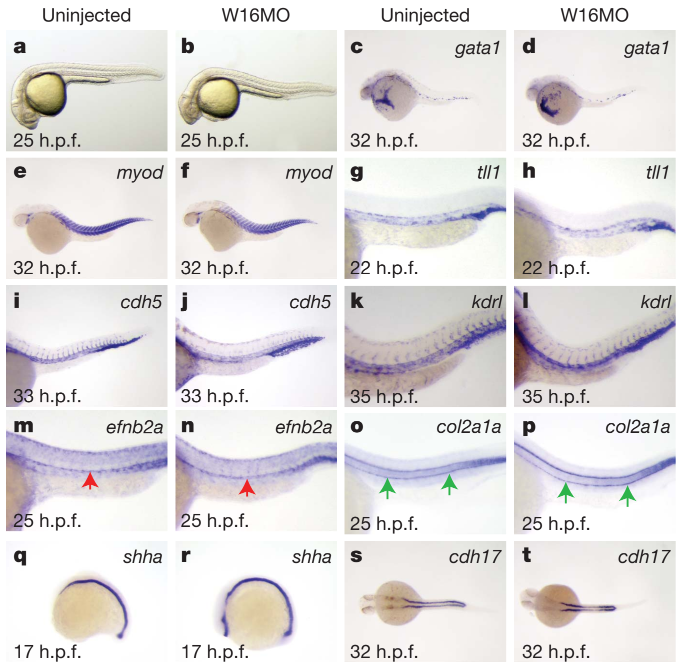


Figure 2 | **The *wnt16* loss-of-function phenotype is specific.** **a**–**t**, Uninjected or W16MO-injected embryos in bright-field (**a**, **b**) or processed by WISH for tissue-specific genes: primitive blood (*gata1*, **c**, **d**), somites (*myod*, **e**, **f**), vasculature (*tll1*, **g**, **h**; *cdh5*, **i**, **j**; *kdrl*, **k**, **l**), dorsal aorta (*efnb2a*, **m**, **n**), hypochord (*col2a1a*, **o**, **p**), floorplate and notochord (*shha*, **q**, **r**), or pronephros (*cdh17*, **s**, **t**) at the developmental times indicated. Red arrows, dorsal aorta (**m**, **n**); green arrows, hypochord (**o**, **p**). Anterior left. **a**–**r**, Dorsal up. **s**, **t**, Dorsal views. **a**–**f**, **q**–**t**, Original magnification at ×64; **g**–**p**, original magnification at ×200.

Supplementary Table 3). Thus, defects in HSC specification in *wnt16* morphant animals are highly specific and not due to wholesale failure in the specification of nearby tissues.

Canonical, β-catenin/Tcf-dependent Wnt signalling has been reported to be involved in HSC specification in mice[2] and very early maintenance in zebrafish[5]. We therefore wanted to determine if Wnt16 is canonical. Overexpression of Wnt16 caused phenotypic defects, demonstrating active protein (Supplementary Fig. 3), but in comparison to Wnt3 (ref. 14) it did not cause ectopic expression of canonical targets (Supplementary Fig. 8). Conversely, Wnt16 knock down caused no decrease in canonical reporter activity (Supplementary Fig. 9). Thus, Wnt16 does not act via β-catenin/Tcf and must act via a non-canonical pathway, as has been suggested for human WNT16B (refs 15, 16). Loss of other non-canonical Wnt ligands causes distinct phenotypes. Loss of Wnt5b is much more severe, causing nearly complete absence of primary trunk vasculature[17], whereas loss of Wnt11 had no effect on HSC specification (Supplementary Fig. 10 and Supplementary Table 4). We therefore conclude that the HSC defects caused by loss of *wnt16* are not a general consequence of loss of non-canonical Wnt signalling.

Notch signalling is required across phyla for the developmental specification of HSCs[1]. Global inhibition of Notch signalling by mutation or targeted deletion of *mind bomb* (*mib*)—which is required for Notch ligand activity—or the essential Notch DNA-binding co-factor *Rbpjκ* results in total loss of HSCs, as well as severe vascular defects[1]. If Wnt16 were to regulate Notch pathway genes, this regulation might explain the loss of HSCs in *wnt16* morphant animals. The specific ligands and receptors required in zebrafish for specification of HSCs by the Notch pathway have not been determined. We therefore examined comparative expression of Notch ligands and receptors that might participate in HSC specification in uninjected or W16MO-injected animals. Most Notch receptors and ligands were either unaffected or very weakly affected (Supplementary Fig. 7 and Supplementary Table 5). However, expression of two Notch ligands, *dlc* and *dld*, was markedly decreased in somites at 17 h.p.f. (16-somite stage (ss); Fig. 3a–d and Supplementary Table 5). In accord with decreased somitic ligand expression, Notch reporter activity was decreased in somites at 17.5 h.p.f. (Supplementary Fig. 11). Interestingly, diminution of *dlc* and *dld* was tissue specific, because expression was maintained in pre-somitic mesoderm (Fig. 4a–d), and *dlc* expression in the dorsal aorta appeared relatively normal at 22 h.p.f. (Supplementary Fig. 12 and Supplementary Table 5). Although Wnt16 function is required for somitic expression of *dlc* and *dld*, we see no evidence for reciprocal regulation of *wnt16* by Dlc and Dld (Supplementary Fig. 13), or indeed by Notch at all, as expression is unaffected in *mib* mutants (not shown).

To determine whether alterations in the expression of *dlc* and/or *dld* might explain the decrease in HSCs we observe in *wnt16* morphant animals, we tested whether loss of function in these genes alone or in combination could alter HSC specification. *Beamter* (*bea*) mutants carry a predicted null allele of *dlc*[18]. We compared HSC and T-lymphocyte marker expression in wild-type embryos, homozygous *bea* mutants, embryos injected with *dld* morpholino (dldMO)[19], and homozygous *bea* mutants injected with dldMO. We found that *runx1* transcripts at 24 h.p.f., and *cmyb* at 36 h.p.f., were greatly reduced in *bea* homozygous mutants and embryos injected with dldMO (Fig. 3e–g, i–k and Supplementary Table 6), but an apparent recovery of HSCs had occurred by 4.5 d.p.f., when *rag1* transcripts (revealing the thymic progeny of HSCs) were readily visible in both groups (Fig. 3m–o and Supplementary Table 6). Combinatorial reduction of Dlc and Dld activity, however, eliminated HSCs, as revealed by elimination of *runx1* transcripts at 24 h.p.f. (Fig. 3h and Supplementary Table 6), *cmyb* transcripts at 36 h.p.f. (Fig. 3l and Supplementary Table 6) and

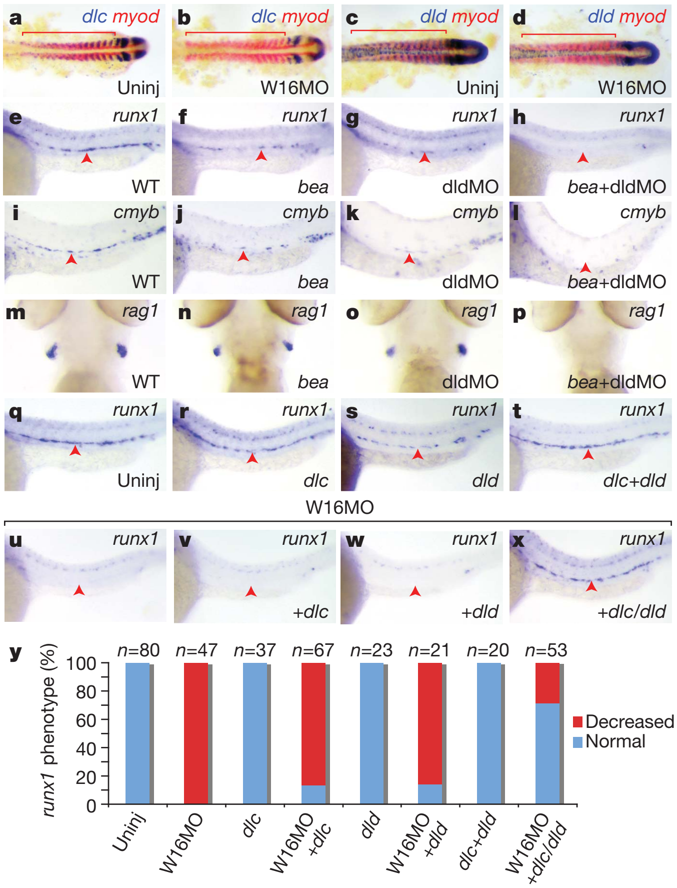


Figure 3 | Wnt16 acts upstream of Notch ligands Dlc and Dld. **a**–**d**, Expression of somitic *dlc* (**a**, **b**) and *dld* (**c**, **d**) but not *myod* (red) is decreased at 17 h.p.f. in W16MO-injected embryos. Red bars indicate somites. **e**–**l**, Expression of the HSC markers *runx1* at 24 h.p.f. (**e**–**h**) and *cmyb* at 36 h.p.f. (**i**–**l**) is reduced in *dlc* mutant (*bea*) embryos (**f**, **j**) or dldMO-injected embryos (**g**, **k**), and eliminated in the combined animals (**h**, **l**). **m**–**p**, The lymphocyte marker *rag1* at 4.5 d.p.f. is present in wild-type (**m**, WT), *bea* (**n**) and dldMO-injected animals (**o**), but eliminated in *bea* embryos injected with dldMO (**p**). **q**–**x**, Combined injection of *dlc* and *dld* rescues HSCs in *wnt16* morphants. *Runx1* at 24 h.p.f. One group of embryos was uninjected, or injected with the indicated Notch ligand mRNAs alone (**q**–**t**). A second group of W16MO-injected embryos was co-injected with Notch ligand mRNAs (**u**–**x**). **y**, Percentages of embryos displaying the depicted phenotypes. **a**–**d**, ×100 (original magnification) views of flat mounts, anterior left. **e**–**l**, **q**–**x**, ×200 (original magnification) lateral views, anterior left, dorsal up. **m**–**p**, ×200 (original magnification) ventral head views, anterior up. Red arrowheads indicate the aorta region.

rag1 transcripts at 4.5 d.p.f. (Fig. 3p and Supplementary Table 6). Thus, *dlc* and *dld* are combinatorially required for specification of HSCs in zebrafish, and their diminished expression in the somites of *wnt16* morphant animals can, in principle, explain the observed loss of HSCs.

To confirm that the loss of HSCs in *wnt16* morphants is due to loss of *dlc* and *dld*, we performed a rescue experiment. We injected embryos with W16MO, and in some cases co-injected mRNA encoding full-length Dlc and Dld ligands singly or in combination. Injection of *dlc* and *dld* mRNA alone or in combination did not have a strong effect on HSC numbers, as measured by $runx1^+$ cells in the dorsal aorta (Fig. 3q–t, y). Co-injection of individual mRNAs with W16MO was unable to restore *runx1* transcript levels (Fig. 3u–w, y). However, when injected together, *dlc* and *dld* restored $runx1^+$ HSCs in a high percentage of W16MO-injected animals (Fig. 3x, y). These results confirm that decreased *dlc* and *dld* expression in *wnt16* morphants is responsible for loss of HSCs.

Studies in zebrafish and mice have shown that Notch specification of HSCs is regulated by Shh and VegfA (ref. 1). Shh regulates the expression of *vegfa*, and VegfA signalling is, in turn, required for the vascular expression of Notch receptor genes[1]. Notch signalling is required for both arteriovenous and HSC specification[1,20–24]. In accord with these observations, loss of Shh signalling causes loss of both artery and HSCs[1]. Taken together, one level of control over HSC specification seems to be through a Shh/VegfA/Notch signalling pathway[1]. Shh/VegfA/Notch specification of HSCs seems to be distinct from Wnt16/Dlc/Dld effects because vascular and arterial specification is unaffected in W16MO animals (Fig. 2i–n and Supplementary Fig. 7a–d, o, p). Moreover, *shha* and Notch receptor expression are unaffected in W16MO animals (Fig. 2q, r and Supplementary Fig. 7a–l). To confirm that these pathways are discrete, we examined the expression of six known Shh target genes, with particular interest in *vegfaa*. None of the Shh targets examined showed significantly altered expression (Supplementary Fig. 14 and Supplementary Table 7). Our results indicate that the Wnt16 and Shh pathways act in parallel upstream of HSC specification.

The predominant model for how Notch signalling regulates HSC specification is that endothelial cells of the dorsal aorta receive a requisite Notch1-mediated signal[1,21–24]. Chimaeric mice generated using Notch1-deleted cells show no contribution of knockout cells to the adult haematopoietic system, demonstrating that Notch1 signalling is required cell-autonomously for specification of HSCs[1,21]. The relevant Notch ligand(s) are thought to be expressed in the formed dorsal aorta and/or immediately surrounding mesenchyme[1,22–24]. A contributing ligand seems to be *Jag1*, because *Jag1*-knockout mice have severely impaired, but not totally abrogated, development of HSCs[1,24]. Notch1 activation of *Runx1* is thought to occur indirectly through *Gata2*, as the *Runx1* promoter does not contain identifiable *Rbpjκ* binding sites[1,21,23,24]. Notch signalling defects in *wnt16* morphants indicate that the requirement for Dlc and Dld is not the same as this cell-autonomous requirement, because decreased Notch ligand expression is specific to somites (Fig. 3a–d), whereas expression of *dlc* is normal in the dorsal aorta of *wnt16* morphants (Supplementary Fig. 12 and Supplementary Table 5) and *dld* is not natively expressed near the formed dorsal aorta. Moreover, expression of *jag1b* and its putative target *gata2* are unaffected in Wnt16 knock down animals (Fig. 4a–d).

To understand better whether Dlc/Dld specification of HSCs works cell-autonomously or not, we sought to define the timing of the Notch signalling event that is absent in W16MO-injected animals using animals carrying transgenes that allow heat-shock induction of the Notch intracellular domain (NICD), which is a dominant activator of the Notch pathway[20,25]. We induced NICD expression at different time points to identify the temporal window when enforced Notch activity could rescue HSC expression in *wnt16* morphants. Nuclear, Myc-tagged NICD protein is present in transgenic animals by 1 h post heat shock, and is strong from 3 h post heat shock (Fig. 4e–h) through to at least 24 h post heat shock (not shown), consistent with previous reports[26]. Heat-shock induction of NICD at 14 h.p.f. (10-ss) rescued *cmyb* expression at 36 h.p.f. (Fig. 4i–k and Supplementary Table 8), as has been seen previously for rescue of *mib*[20]. Surprisingly, heat shock just 2 h later (14-ss) did not rescue *cmyb* expression along the dorsal aorta (red circle, Fig. 4l and Supplementary Table 8). These results indicate that the critical phase of Notch signalling required for HSC specification downstream of Wnt16 occurs between 15–17 h.p.f. (12–16-ss) and abruptly terminates by about 18–19 h.p.f. (18–20-ss).

Because this timing is well before the formation of the dorsal aorta and HSCs from its ventral endothelium, we sought to determine when cells fated to become HSCs first experience a cell-autonomous Notch signal. First, we used a *tp1:Kaede* Notch reporter line expressing a green-to-red, photoconvertible Kaede protein under the control of a Notch-responsive promoter. Photoconversion of Notch-responsive, $Kaede^+$ cells in the Dlc/Dld-critical window, before 19.5 h.p.f. (21-ss), yielded converted, red fluorescent progeny that contributed to the region near the dorsal aorta at 3 d.p.f. (left panels of Fig. 4m, n), but these cells never became HSCs, as shown by their failure to produce

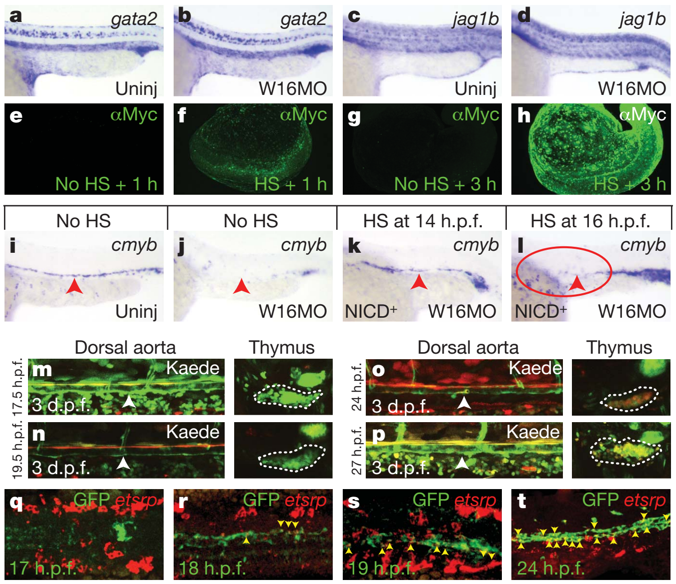


Figure 4 | Non-cell-autonomous requirement for Notch in HSC specification. **a**–**d**, Expression of the Notch target *gata2* (**a**, **b**) and the Notch ligand *jag1b* (**c**, **d**) in uninjected (**a**, **c**) or W16MO-injected (**b**, **d**) embryos at 22 h.p.f. **e**–**h**, Whole-mount immunofluorescence visualization of the Myc-tagged NICD at 1 h (**e**, **f**) and 3 h (**g**, **h**) after either no induction (**e**, **g**) or heat-shock (HS) induction (**f**, **h**). **i**–**l**, *Cmyb* expression at 36 h.p.f. in transgenic animals carrying a heat-shock-inducible dominant activator of Notch signalling (NICD) in uninjected (**i**), W16MO-injected (**j**), W16MO-injected and heat-shock induced at 14 h.p.f. (10-ss; **k**) or 16 h.p.f. (14-ss; l). Red arrowheads indicate the dorsal aorta region. Red circle indicates the area where HSCs should normally be expressed (**l**). **m**–**p**, Green-to-red photoconvertible Kaede Notch reporter animals were entirely photoconverted at the times indicated at the left of each panel pair and imaged at 3 d.p.f. Confocal images of the dorsal aorta (white arrowheads; left panels of **m**–**p**) and thymus (dashed white outline; right panels of **m**–**p**) reveal photoconverted cells only in the thymi of fish converted at 24 and 27 h.p.f. (right panels of **o**, **p**). **q**–**t**, Max-projection confocal images of the trunk region of embryos processed by double fluorescence *in situ* for a Notch reporter *GFP* transgene (green) and the haematopoietic mesoderm marker *etsrp* (red) at the times indicated. Yellow arrowheads indicate double-positive cells (**r**–**t**). **a**–**d**, **i**–**l**, ×200 (original magnification) lateral views of the trunk region, anterior left, dorsal up. **e**–**h**, ×100 (original magnification) whole-embryo views. **m**–**p**, Cropped, ×200 (original magnification) lateral views of the dorsal aorta (left panels). **m**–**p**, Single thymic lobes (right panels). **q**–**s**, Cropped, ×200 (original magnification) dorsal views, anterior left. **t**, Cropped, ×100 (original magnification) lateral aorta view, anterior left, dorsal up.

labelled thymic progeny (right panels of Fig. 4m, n). In contrast, photoconversion at 24 h.p.f. and later time points produced both labelled dorsal aorta (left panels of Fig. 4o, p) and thymic immigrants (right panels of Fig. 4o, p). Because there is a lag between reception of a Notch signal and the production of mature Kaede protein in the reporter lines, we wanted to confirm when Notch activity appears in pre-haematopoietic cells by a more immediate readout for Notch responsiveness. We performed double fluorescence *in situ* analysis in *tp1:GFP* Notch-reporter transgenic animals for expression of *GFP* and *etsrp*, which labels pre-haematopoietic mesoderm and is required for HSC specification[27]. The very first double-labelled cells appeared at 18 h.p.f. in the midline (Fig. 4q, r), and increased in number through to at least 24 h.p.f. (Fig. 4s, t). Taken together, our data indicate that Notch signalling in cells fated to become HSCs begins after 18 h.p.f. and continues well into the second day of development. This timing fits well with the established model of cell-autonomous Notch specification of HSCs, where Notch signalling in nascent HSCs occurs after aortic specification via interactions between Jag1b and Notch1 (refs 1, 21–24). Overall, our results indicate that somitic signalling by Dlc and Dld downstream of Wnt16 is temporally and spatially distinct from the observable cell-autonomous requirement for Notch signalling. Wnt16 therefore controls a previously unappreciated environmental requirement for Notch signalling in the somites, most simply explained by regulation of an unidentified relay signal.

To understand better what cell population Wnt16-regulated somitic Dlc and Dld act on, we examined fine somite patterning in Wnt16 and Dlc/Dld loss-of-function animals. Because somite segmentation (Figs 2a, b, e, f and 3a–d) and myotomal specification (Figs 2e, f and 3a–d) appear to be intact in *wnt16* morphants, we examined specification of the sclerotomal somite compartment, which houses vertebral and vascular smooth muscle cell precursors. Sclerotomal markers displayed severely decreased expression, albeit to variable extents, in both W16MO and combined *dlc*/*dld* loss of function animals (Supplementary Fig. 15 and Supplementary Table 9). Although hypochord (Supplementary Fig. 15u–y, green arrowheads) was abolished in the *dlc*/*dld* double loss-of-function animals (Supplementary Fig. 15x), as has been reported[18], hypochord was specified normally in *wnt16* morphants (Fig. 2o, p and Supplementary Fig. 15y), emphasizing the fact that the Wnt16-dependent loss of somitic *dlc*/*dld* is distinct from global loss. Our results indicate that sclerotome specification or morphogenesis is required for HSC specification.

Our results demonstrate that non-canonical signalling by Wnt16 is required genetically upstream of the combined actions of the Notch pathway ligands Dlc and Dld for HSC specification. Dlc/Dld-mediated Notch signalling is spatially and temporally distinct from previously described cell-autonomous requirements for Notch in HSC specification, pointing to the possibility of a novel relay signal. These data represent the first demonstration, to our knowledge, that non-canonical Wnt signalling activity is required for HSC specification in vertebrates. Given that *Wnt16* in mouse is expressed at similar times of development[28] and is expressed in embryoid bodies during commitment to blood and vasculature[29], it is feasible that this function is conserved in mammals.

METHODS SUMMARY

Zebrafish strains. The following strains were maintained in accordance with IACUC approved procedures: AB*, *Tg(TOP:GFP)*^{w25}, *Tg(-6.0itga2b:eGFP)*^{la2}, *Tg(hsp70l:Gal4)*$^{1.5kca4}$, *Tg(UAS:myc-Notch1a-intra)*^{kca3}, *Tg(tp1-MmHbb:EGFP)*^{um14}, *Tg(tp1-MmHbb:Kaede)*^{um15}, *Tg(cmyb:EGFP)*^{zf169}, *Tg(kdrl:RFP)*^{la4} and *dlc*$^{tit446/tit446}$.

Constructs. Probes and mRNA were synthesized from published constructs or cloned *de novo* according to standard procedures.

Morpholinos. The following morpholinos were used: 5 ng W16MO1 (AGGTTAGTTCTGTCACCCACCTGTC), W16MO2 (GCGTGGAATACTTACATCCAACTTC) and W16CoMO (AcGTTAGTTgTGTCAgCCAgCTcTC; lowercase letters denote mismatched bases).

Genotyping and PCR. Genotyping and PCR were performed by standard methods using primers and conditions described in Methods.

Whole-mount *in situ*, antibody staining and microscopy. These were performed according to standard methods and are described in full in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 May 2010; accepted 11 April 2011.

1. Gering, M. & Patient, R. Notch signalling and haematopoietic stem cell formation during embryogenesis. *J. Cell. Physiol.* **222,** 11–16 (2010).
2. Staal, F. J. & Luis, T. C. Wnt signaling in hematopoiesis: crucial factors for self-renewal, proliferation, and cell fate decisions. *J. Cell. Biochem.* **109,** 844–849 (2010).
3. Angers, S. & Moon, R. T. Proximal events in Wnt signal transduction. *Nature Rev. Mol. Cell Biol.* **10,** 468–477 (2009).
4. Takada, S. *et al.* Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev.* **8,** 174–189 (1994).
5. Goessling, W. *et al.* Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration. *Cell* **136,** 1136–1147 (2009).
6. McWhirter, J. R. *et al.* Oncogenic homeodomain transcription factor E2A-Pbx1 activates a novel WNT gene in pre-B acute lymphoblastoid leukemia. *Proc. Natl Acad. Sci. USA* **96,** 11464–11469 (1999).
7. Bertrand, J. Y. *et al.* Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464,** 108–111 (2010).

8. Kissa, K. & Herbomel, P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* **464,** 112–115 (2010).
9. Boisset, J. C. *et al. In vivo* imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* **464,** 116–120 (2010).
10. Bertrand, J. Y. *et al.* Definitive hematopoiesis initiates through a committed erythromyeloid precursor in the zebrafish embryo. *Development* **134,** 4147–4156 (2007).
11. Kissa, K. *et al.* Live imaging of emerging hematopoietic stem cells and early thymus colonization. *Blood* **111,** 1147–1156 (2008).
12. Lin, H. F. *et al.* Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. *Blood* **106,** 3803–3810 (2005).
13. Yokota, T. *et al.* Tracing the first waves of lymphopoiesis in mice. *Development* **133,** 2041–2051 (2006).
14. Clements, W. K., Ong, K. G. & Traver, D. Zebrafish wnt3 is expressed in developing neural tissue. *Dev. Dyn.* **238,** 1788–1795 (2009).
15. Nygren, M. K. *et al.* β-catenin is involved in N-cadherin-dependent adhesion, but not in canonical Wnt signaling in E2A–PBX1-positive B acute lymphoblastic leukemia cells. *Exp. Hematol.* **37,** 225–233 (2009).
16. Teh, M. T. *et al.* Role for WNT16B in human epidermal keratinocyte proliferation and differentiation. *J. Cell Sci.* **120,** 330–339 (2007).
17. Cirone, P. *et al.* A role for planar cell polarity signaling in angiogenesis. *Angiogenesis* **11,** 347–360 (2008).
18. Julich, D. *et al. beamter/deltaC* and the role of Notch ligands in the zebrafish somite segmentation, hindbrain neurogenesis and hypochord differentiation. *Dev. Biol.* **286,** 391–404 (2005).
19. Holley, S. A., Julich, D., Rauch, G. J., Geisler, R. & Nusslein-Volhard, C. *her1* and the *notch* pathway function within the oscillator mechanism that regulates zebrafish somitogenesis. *Development* **129,** 1175–1183 (2002).
20. Burns, C. E., Traver, D., Mayhall, E., Shepard, J. L. & Zon, L. I. Hematopoietic stem cell fate is established by the Notch-Runx pathway. *Genes Dev.* **19,** 2331–2342 (2005).
21. Hadland, B. K. *et al.* A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development. *Blood* **104,** 3097–3105 (2004).
22. Kumano, K. *et al.* Notch1 but not Notch2 is essential for generating hematopoietic stem cells from endothelial cells. *Immunity* **18,** 699–711 (2003).
23. Robert-Moreno, A., Espinosa, L., de la Pompa, J. L. & Bigas, A. RBPjκ-dependent Notch function regulates *Gata2* and is essential for the formation of intra-embryonic hematopoietic cells. *Development* **132,** 1117–1126 (2005).
24. Robert-Moreno, A. *et al.* Impaired embryonic haematopoiesis yet normal arterial development in the absence of the Notch ligand Jagged1. *EMBO J.* **27,** 1886–1895 (2008).
25. Scheer, N. & Campos-Ortega, J. A. Use of the Gal4-UAS technique for targeted gene expression in the zebrafish. *Mech. Dev.* **80,** 153–158 (1999).
26. Scheer, N., Groth, A., Hans, S. & Campos-Ortega, J. A. An instructive function for Notch in promoting gliogenesis in the zebrafish retina. *Development* **128,** 1099–1107 (2001).
27. Ren, X., Gomez, G. A., Zhang, B. & Lin, S. *Scl* isoforms act downstream of *etsrp* to specify angioblasts and definitive hematopoietic stem cells. *Blood* **115,** 5338–5346 (2010).
28. Kemp, C., Willems, E., Abdo, S., Lambiv, L. & Leyns, L. Expression of all Wnt genes and their secreted antagonists during mouse blastocyst and postimplantation development. *Dev. Dyn.* **233,** 1064–1075 (2005).
29. Corrigan, P. M., Dobbin, E., Freeburn, R. W. & Wheadon, H. Patterns of Wnt/Fzd/LRP gene expression during embryonic hematopoiesis. *Stem Cells Dev.* **18,** 759–772 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors wish to thank L. Zon, K. Poss, D. Kimelman, M. Lardelli, B. Appel, C. Burns, J. Campos-Ortega, D. Ransom, N. Trede, J. Lewis, M. Pack, S. Holley, C. Moens, B. Paw, R. Karlström and J. Waxman for probe constructs. L. Zon, R. Dorsky, S. Lin and S. Holley provided transgenic and mutant zebrafish lines. C. Weaver, K. Willert, K. J. P. Griffin, J. Bertrand, D. Stachura and Y. Lee provided critical evaluation of the manuscript. This research was funded by an AHA Postdoctoral Fellowship 0725086Y to W.K.C., an AHA Predoctoral Founders Affiliate Fellowship 0815732D to J.C.M., NIH R01-HL093467 to N.L. and NIH R01-DK074482, CIRM New Investigator Award, and March of Dimes 6-FY09-508 to D.T.

Author Contributions W.K.C. and D.T. designed all experiments. Whole-mount immunofluorescence, double fluorescence *in situs*, and Kaede-based fate mapping was performed by A.D.K. K.G.O. cloned and subcloned multiple constructs. J.C.M and N.L. generated Notch reporter lines. All other experiments were performed by W.K.C. The manuscript was written by W.K.C. and edited by N.L. and D.T., with critical input as described in the Acknowledgments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.T. (dtraver@ucsd.edu).

METHODS

Zebrafish husbandry, microinjections and heat shock. Zebrafish strains AB*, *Tg(TOP:GFP)*^{w25}, *Tg(-6.0itga2b:eGFP)*^{la2}, *Tg(hsp70l:Gal4)*$^{1.5kca4}$, *Tg(UAS:myc-Notch1a-intra)*^{kca3}, *Tg(tp1-MmHbb:EGFP)*^{um14}, *Tg(tp1-MmHbb:Kaede)*^{um15}, *Tg(cmyb:EGFP)*^{zf169}, *Tg(kdrl:RFP)*^{la4} and *dlc*$^{tit446/tit446}$ (S. Holley) were maintained, crossed, injected, raised and staged as described[30], and in accordance with IACUC guidelines. The new line *Tg(tp1-MmHbb:Kaede)*^{um15} was generated by Tol2-mediated transgenic insertion of a transgene containing a previously described Notch-responsive promoter[31] driving *Kaede* expression. Heat shocks were performed at the times indicated for 45 min at 37 °C, as previously described[20].

Cloning, constructs and probes. Both zebrafish *wnt16* isoforms (*wnt16-001* and *wnt16-002*, Zebrafish Information Network (ZFIN)) were amplified from AB* embryo cDNA at tailbud, 9-somite and 24 h.p.f., using primers (5′ UTR-Wnt16-F CAGGTGCTACATATTAGATGCAGTGG, Wnt16-var.-F GACATGG ATAATACCGGTTGTGGG, and zWnt16-R TTACTTGCAGGTGTGCATGT CATTC) designed based on GenBank sequences NM_207096 and CD751181.1 and cloned to pCRII-TOPO-TA (Invitrogen) according to the manufacturer's instructions. Sequenced clones conform to subsequently deposited GenBank sequences NM_001100046.1 (*wnt16-001*) and BC066432.1 (*wnt16-002*), with no non-silent alterations. The *wnt16-001* form was re-cloned using the primers ggatccGACATGGATAATACCGGTTGTG and ctcgagTTACTTGCAGGTGTG CATGTC to introduce 5′ BamH1 and 3′ Xho1 sites (PCR-added cut sites presented as lowercase in all primer sequences). *Wnt16-001* and *wnt16-002* sequences were subcloned to pCS2+ (ref. 32) via BamH1/Xho1 or EcoR1, respectively. Full-length *dlc* was amplified from 16-ss AB* cDNA using the primers dlc-F (ctcgagAA GATGGCTCGTGTTTTATTAAC) and dlc-R (tctagaCTATACCTCAGTAGC AAACACACG), TOPO-TA cloned, confirmed by sequencing, and subcloned to pCS2+. pCS2+ *wnt3*, pCS2+ *dnfgfr1-eGfp* (K. Poss) and pBS *chd* (D. Kimelman) were described previously[14]. The following probe and expression constructs were gifts as indicated: pCS2+ *runx1* (C. Burns), pBK-CMV *scl* (L. Zon), pBK-CMV *cmyb* (L. Zon), pBS *kdrl* (D. Ransom), pBS *gata1* (D. Ransom), pCS2+ *gata2* (B. Paw), pCRII *rag1* (N. Trede), pBS *dlc* (J. Lewis), pBS *dld* (B. Appel), pCS2+ *dld* (S. Holley), pBS *jag1b* (M. Pack), pBS *jag2* (M. Pack), pCR-Script *notch1a* (J. Campos-Ortega), pCR-Script *notch1b* (M. Lardelli), pCR-Script *notch2* (B. Appel), pCR-Script *notch3* (M. Lardelli), pSPORT1 *etsrp* (S. Sumanas), pBS *nkx2.2* (R. Karlström), pCRII *gli1* (R. Karlström), pBS *ptc2* (R. Karlström), pBS *ptc1* (J. Waxman), pBS *prdm1a* (J. Waxman). The following probe constructs were amplified *de novo* and cloned to pCRII-TOPO-TA (Invitrogen) using primers as indicated: pCRII *myod* (myod-F AAGATGGAGTTGTCGGATATCC, myod-R AGAATTTTAAAGCACTTGATAAATGG), pCRII *cdh5-frag* (cdh5-probe-F TGCCTCCGACAAGGATGAAA, cdh5-probe-R ACCGAGGTCCCCACTCAT GT), pCRII *cdh17/cb903* (cb903/cdh17-F GCGGATGATACAGGAACAGG, cb903/cdh17-R CTGAAGGCAGATGAAGCCC), pCRII *col2a1a-frag* (col2a1a-probe-F CCACCTGGATTGACTGGACC, col2a1a-probe-R GTAGTGCTTGCA TGTTCGGTC), pCRII *vegfaa*$_{165}$ (vegfAa-long-F GTTAATTTTAGCGGATTCG ACG, vegfAa-short-R GATCATCATCTTGGCTTTTCAC), pCRII *shha* (Bam-shha-F ggatccAAAATGCGGCTTTTGAC, shha-R1-R gaattcTCAGCTTGAGTT TACTGACATCC; subsequently subcloned to pCS2+), pCRII *foxc1a* (foxc1a-F GTCATGCAGGCGCGCTATT, foxc1a-R ctcgagTCAAAATTTGCTGCAGTCA TACAC), pCRII *foxc1b* (foxc1b-F ggatccACGATGCAGGCGCGCTACCC, foxc1b-R TCAGAACTTGCTGCAGTCGTATAC), pCRII *twist1b* (twist1b-F GAGATGCCCGAAGAGCCCGCGC, twist1b-R ctcgagCTAGTGAGATGCAGA CATGGACC), pCRII *twist2* (twist2-F GAAATGGAAGAGAGTTCTAGCTC, twist2-R ctcgagCTAGTGGGACGCAGACATCG). A *pax1* fragment corresponding exactly to the 1,080-bp open reading frame found in NM_001080592, which has been annotated as the zebrafish *Pax1* orthologue at NCBI Homologene (HomoloGene:4514), was amplified (R1-pax1-F gaattcAAAATGCTTTCGTGT TTTGCAGAG, pax1-Xba-R tctagaTTACGAGGATGAGGTAGAAAGGC) from 24 h.p.f. AB* cDNA to generate pCRII *pax1*. The *pax1* gene is located on chromosome 17 and shows syntenic conservation of the 5′ neighbour (*Nkx2.2* in mouse and *NKX2.2* in human). The encoded protein is 69% identical and 90% similar to mouse and human PAX1. A 5′ Xho1/Cla1 fragment of pBS *efnb2a* (gift of C. Moens) was re-cloned to pBS to generate pBS *efnb2a-probe*. Digoxigenin- and fluorescein-labelled probes were generated as described previously[33] using 5′ cut sites and RNA polymerases as appropriate (details available on request). Embryos were mounted and photographed as described[14].

mRNA, morpholinos and injections. 5′-G-capped mRNAs were synthesized from Not1- or Asp718-(pCS2+ dlc) linearized pCS2+ constructs as described previously[14], using the mMessage mMachinge kit (Ambion). The following morpholino antisense oligonucleotides were synthesized by Gene Tools, LLC and suspended as 25 mg ml^{-1} stocks in DEPC ddH_2O and diluted to injection strengths: W16MO1 AGGTTAGTTCTGTCACCCACCTGTC, W16MO2 GCGT GGAATACTTACATCCAACTTC, W16CoMO2 ('CoMO' in the text) AcGTT AGTTgTGTCAgCCAgCTcTC, lowercase letters denote mismatched bases, dldMO2 AAACAGCTATCATTAGTCGTCCCAT (ref. 19), W11MO GAAAGT TCCTGTATTCTGTCATGTC (ref. 34). Injections were performed as described previously[14], and 0.1% phenol red (Sigma Aldrich) was included as an indicator. W16MO1 and W16MO2 were injected individually at 5 ng and combinatorially at 2 ng W16MO1 plus 3 ng W16MO2. In all depicted cases, both morpholinos caused the representative phenotypes shown, thus embryos are labelled 'W16MO'. A total of 5 or 7 ng of dldMO and W11MO were injected. 50 pg of *dlc* and/or *dld* mRNA were injected for rescue experiments.

Genomic and phylogenetic analyses. Alignments, genomic analyses and phylogenetic comparisons were performed as described previously[14] using the following sequences: Wnt16l NP_001093516.1, mWnt16 NP_444346.3, hWNT16a NP_057171.2, hWNT16b NP_476509.1 and hWNT4 NP_110388.2 (as the outgroup). Sequencing analysis was performed with Sequencher software (GeneCodes Corp.).

PCR genotyping and RT–PCR. Fixed, WISH-processed individual embryos had DNA isolated in lysis buffer (10 mM Tris, pH 8.3, 50 mM KCl, 0.3% Tween-20, 0.3% NP-40), 98 °C, 10 min, held at 4 °C to allow addition of proteinase K to a final concentration of 1 mg ml^{-1}, 18 h at 55 °C, 15 min at 98 °C. Presence of the *UAS:NICD* transgene was assessed by PCR using the primers E1B-F CATCGCGTCTCAGCCTCAC and Notch-R CGGAATCGTTTATTGGTGT CG (T_m 55 °C, extension time 45 s, 35 cycles), with ef-1α-F GTGCTG TGCTGATTGTTGCT, ef-1α-R TGTATGCGCTGACTTCCTTG as a positive control (T_m 56 °C, extension time 30 s, 25 cycles). For RT–PCR, RNA was isolated from groups of 30 whole embryos at the stages indicated, and cDNA prepared as described previously[14]. PCR on cDNA was amplified using the primers *ef-1α* (as above), wnt16-RT1-F ACTAAAGAGACAGCGTTCATCC, wnt16-RT1-R AACT CATCTTTGGTGATAGGC, wnt16-RT3-F TTGTGGGATACATGCAGTTCA and wnt16-RT3-R CACAGCTCCTTCTGCTTGTG with Taq polymerase (Invitrogen) at a T_m of 56 °C, extension times of 30 s, 38 cycles. Gels were imaged as described previously[14].

Whole-mount *in situ* and antibody staining. Single and double enzymatic WISH was performed as described previously[33]. Double fluorescence *in situs* were performed according to published protocols[18]. Whole-mount immunofluorescence was performed as described[35], using anti-Myc monoclonal 9E10 antibodies at 1:200 (Covance) and Dylight488 AffiniPure donkey anti-mouse IgG secondary antibodies (Jackson Immunoresearch Laboratories) at 1:100.

Confocal, fluorescence microscopy, photoconversion and time-lapse imaging. Confocal images were acquired, essentially as described[7]. Photoconversion of Kaede proteins was performed using the Leica SP5 ROI and bleach functions on whole embryos, excluding heads, using a 405-nm laser line for 2 min. Kaede fluorescence was visualized using 488-nm and 543-nm laser lines. Fluorescence images of transgenic animals were acquired on a Leica DMI 6000 (Leica Microsystem) and time-lapse images were analysed and processed using Volocity software (Perkin-Elmer) as described previously[36].

30. Westerfield, M. The Zebrafish Book. *A Guide for the Laboratory Use of Zebrafish (*Danio Rerio*)* (Univ. Oregon Press, 2004).
31. Parsons, M. J. *et al.* Notch-responsive cells initiate the secondary transition in larval zebrafish pancreas. *Mech. Dev.* **126,** 898–912 (2009).
32. Turner, D. L. & Weintraub, H. Expression of achaete-scute homolog 3 in *Xenopus* embryos converts ectodermal cells to a neural fate. *Genes Dev.* **8,** 1434–1447 (1994).
33. Clements, W. K. & Kimelman, D. LZIC regulates neuronal survival during zebrafish development. *Dev. Biol.* **283,** 322–334 (2005).
34. Lele, Z., Bakkers, J. & Hammerschmidt, M. Morpholino phenocopies of the *swirl*, *snailhouse*, *somitabun*, *minifin*, *silberblick*, and *pipetail* mutations. *Genesis* **30,** 190–194 (2001).
35. Nüsslein-Volhard, C. & Dahm, R. *Zebrafish* (Oxford Univ. Press, 2002).
36. Bertrand, J. Y., Kim, A. D., Teng, S. & Traver, D. CD41$^+$ cmyb$^+$ precursors colonize the zebrafish pronephros by a novel migration route to initiate adult hematopoiesis. *Development* **135,** 1853–1862 (2008).

doi:10.1038/nature10106

Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1

Momoko Maekawa[1,2], Kei Yamaguchi[3], Tomonori Nakamura[1,4], Ran Shibukawa[1,2], Ikumi Kodanaka[1,2], Tomoko Ichisaka[1,4], Yoshifumi Kawamura[3], Hiromi Mochizuki[3], Naoki Goshima[5] & Shinya Yamanaka[1,2,4,6]

Induced pluripotent stem cells (iPSCs) are generated from somatic cells by the transgenic expression of three transcription factors collectively called OSK: Oct3/4 (also called Pou5f1), Sox2 and Klf4[1]. However, the conversion to iPSCs is inefficient. The proto-oncogene *Myc* enhances the efficiency of iPSC generation by OSK but it also increases the tumorigenicity of the resulting iPSCs[2]. Here we show that the Gli-like transcription factor Glis1 (Glis family zinc finger 1) markedly enhances the generation of iPSCs from both mouse and human fibroblasts when it is expressed together with OSK. Mouse iPSCs generated using this combination of transcription factors can form germline-competent chimaeras. Glis1 is enriched in unfertilized oocytes and in embryos at the one-cell stage. DNA microarray analyses show that Glis1 promotes multiple pro-reprogramming pathways, including Myc, Nanog, Lin28, Wnt, Essrb and the mesenchymal–epithelial transition. These results therefore show that Glis1 effectively promotes the direct reprogramming of somatic cells during iPSC generation.

The generation of iPSCs is technically simple and highly reproducible[3,4] but only a small proportion of cells become iPSCs after introduction of the four transcription factors[5]. In addition, the generation of iPSCs is slow and requires multiple cell divisions[6]. Reprogramming towards pluripotency can also be achieved by nuclear transfer to meiotic oocytes[7] or mitotic zygotes[8]: this strategy is technically more demanding but it is efficient, rapid and independent of cell division. These differences may indicate that oocytes and zygotes contain factor(s) that promote reprogramming during the generation of iPSCs.

In this study, we initially evaluated a library of 1,437 human transcription factors for their ability to replace Kruppel-like factor 4 (Klf4) or POU domain, class 5, transcription factor 1 (Pou5f1, also known as Oct3/4) during iPSC generation from mouse skin fibroblasts containing a green fluorescent protein (GFP) reporter driven by the nanog homeobox (*Nanog*) promoter and enhancers[9] (Supplementary Table 1). We found that 18 factors could replace Klf4 reproducibly, although with much lower efficiencies of iPSC generation (Supplementary Table 2); we failed to identify any factors that replaced Oct3/4.

Among these 18 factors, we found that GLIS1, a GLI transcription factor[10], markedly increased the number of GFP-positive colonies when it was co-introduced with the 'OSK' transcription factors Oct3/4, SRY-box 2 (Sox2) and Klf4 into adult mouse skin fibroblasts (Fig. 1a). The effect of GLIS1 was comparable to that of MYC, as judged by the number of GFP-positive colonies (Fig. 1b). We also observed a synergistic increase in the number of GFP-positive colonies when both GLIS1 and MYC were co-introduced with OSK. Notably, GLIS1 specifically promoted the generation of GFP-positive colonies, but not GFP-negative colonies, which represent either partially reprogrammed cells or transformed cells (Fig. 1c). In contrast, MYC increased the number of GFP-negative colonies more than the number of GFP-positive ones. This undesired effect of MYC was counteracted when GLIS1 was co-expressed.

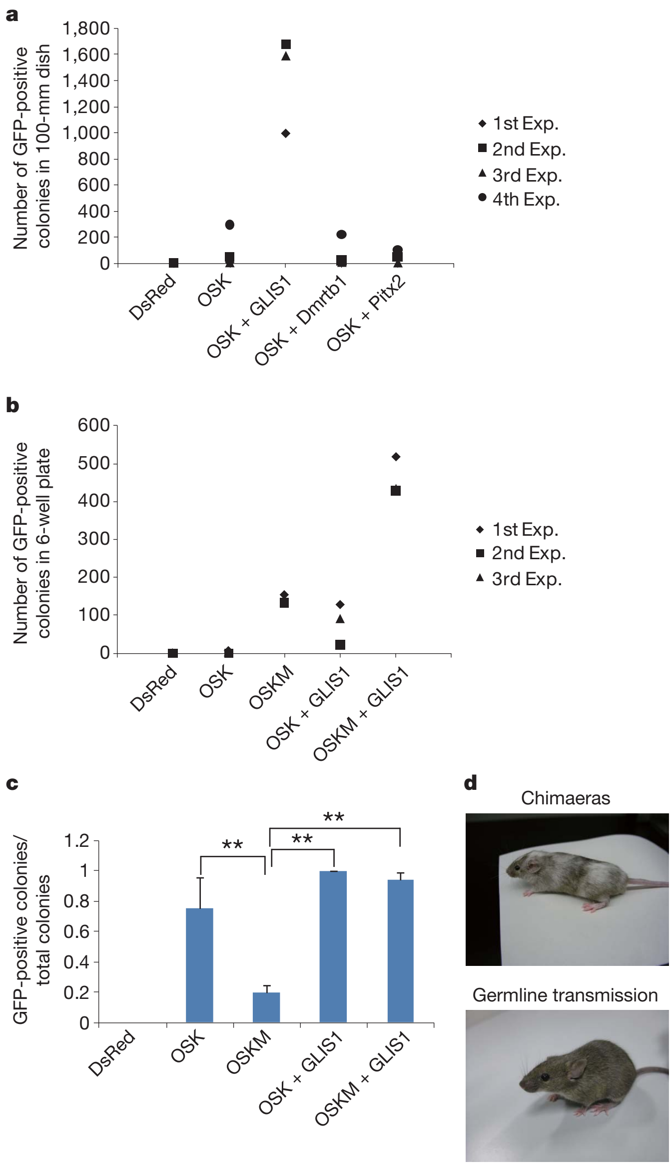


Figure 1 | Promotion of mouse iPSC generation by GLIS1. **a**, Number of *Nanog*–GFP-positive colonies from mouse skin fibroblasts in a 100-mm dish, 28 d after infection. Three days after infection, fibroblasts were re-seeded on feeder cells. Exp, experiment. **b**, Number of *Nanog*–GFP-positive colonies from mouse skin fibroblasts in a 6-well plate, 22 d after infection. **c**, Proportion of *Nanog*–GFP-positive colonies to total number of colonies (derived directly from experiments in **b**). Error bars, s.d.; $n = 3$; **, $P < 0.01$. **d**, Upper panel: chimaeric mouse derived from iPSCs obtained by transfection of MEFs with OSK + GLIS1. Lower panel: coat colour of offspring, showing germline transmission.

[1]Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto 606-8507, Japan. [2]Yamanaka iPS Cell Special Project, JST, Kawaguchi 332-0012, Japan. [3]Japan Biological Informatics Consortium, Tokyo 135-0064, Japan. [4]Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto 606-8507, Japan. [5]Biomedicinal Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan. [6]Gladstone Institute of Cardiovascular Disease, San Francisco, California 94158, USA.

Mouse iPSCs generated with OSK and GLIS1 showed morphologies similar to embryonic stem (ES) cells (Supplementary Fig. 1a). Pluripotency markers such as *Nanog* were expressed at comparable levels to those in ES cells (Supplementary Fig. 1b) and the iPSCs formed teratomas in nude mice (Supplementary Fig. 1c). Furthermore, they produced germline-competent chimaeras (Fig. 1d and Supplementary Table 3).

In human adult fibroblasts, GLIS1 showed a similar effect: it promoted the generation of ES-cell-like colonies to a comparable degree to MYC when it was co-introduced with OSK (Fig. 2a). Notably, GLIS1 specifically promoted the generation of ES-cell-like colonies with a flat, round shape and a distinct edge, but did not promote the generation of non-ES-cell-like colonies, which were granular with an irregular edge (Fig. 2b and Supplementary Fig. 2). In contrast, MYC increased the number of non-ES-cell-like colonies more than the number of ES-cell-like ones (Fig. 2b). The iPSCs generated with OSK and GLIS1 were similar to ES cells in morphology (Supplementary Fig. 3a) and in their expression of undifferentiated-ES-cell marker genes, such as *OCT3/4, SOX2, NANOG* and *ZFP42* (zinc finger protein 42 homolog (mouse), also known as *REX1*) (Supplementary Fig. 3b). DNA microarray analyses showed that human iPSCs established with OSK and GLIS1 had similar global gene expression to cells generated with OSK and MYC (OSKM) (Fig. 2c). The promoter region of the *OCT3/4* gene showed a hypomethylation pattern (Supplementary Fig. 3c) and the iPSCs differentiated into various cells of the three germ layers in the embryoid body (Supplementary Fig. 3d) and also into teratomas (Fig. 2d). These results demonstrate that GLIS1 strongly and specifically promotes the generation of both mouse and human iPSCs by OSK.

We next studied the expression pattern of *Glis1* in mouse cells. Analyses of expressed sequence tag (EST) databases predicted that *Glis1* expression would be enriched in zygotes, especially in the fertilized ovum (http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist=Mm.331757 as of 7 December 2010). In addition, the gene expression data from reverse transcription PCR (RT–PCR), provided by the mouse genome database MGI, showed that there was moderate expression of *Glis1* in metaphase II oocytes and weak expression in two-cell embryos, but that expression was either absent or at trace levels in embryos at the four-cell to embryonic-day-4.5 stages (http://www.informatics.jax.org/searches/expression.cgi?32989 as of 7 December 2010, also reported in ref. 11 in their Supplementary Table 1). To confirm the specific expression of *Glis1* in oocytes and one-cell embryos, we isolated total RNA from oocytes, early embryos and several adult mouse tissues. Real-time PCR detected the highest expression of *Glis1* in one-cell embryos and unfertilized eggs. A modest level of expression was detected in two-cell embryos and placentas and weak expression was detected in several adult tissues (Fig. 3a). These data confirmed that *Glis1* RNA is enriched in unfertilized eggs and one-cell embryos.

We next examined whether endogenous Glis1 has a role during iPSC generation by OSK. We found that Glis1 is expressed at a low level in mouse fibroblasts before and after the introduction of OSK (Supplementary Fig. 4a). We constructed retroviral vectors to express several *Glis1* small hairpin RNAs (shRNAs), as well as scrambled controls, and tested the knockdown efficiency of each shRNA retrovirus in skin fibroblasts. We found that shRNA2 and shRNA6 were effective (Supplementary Fig. 4b). We then introduced each of these shRNAs, together with OSK, into mouse embryonic fibroblasts (MEFs) containing the *Nanog*–GFP reporter. We found that both shRNA2 and shRNA6 significantly decreased the number of GFP-positive colonies (Supplementary Fig. 4c), in contrast to the scrambled control shRNA. These results show that endogenous Glis1 may have a supportive role during the generation of mouse iPSCs by OSK.

Finally, we tried to elucidate how Glis1 enhances iPSC generation by OSK. We previously reported that suppression of the p53 pathway markedly enhanced iPSC generation from both mouse and human cells[12]. We therefore hypothesized that Glis1 may enhance direct

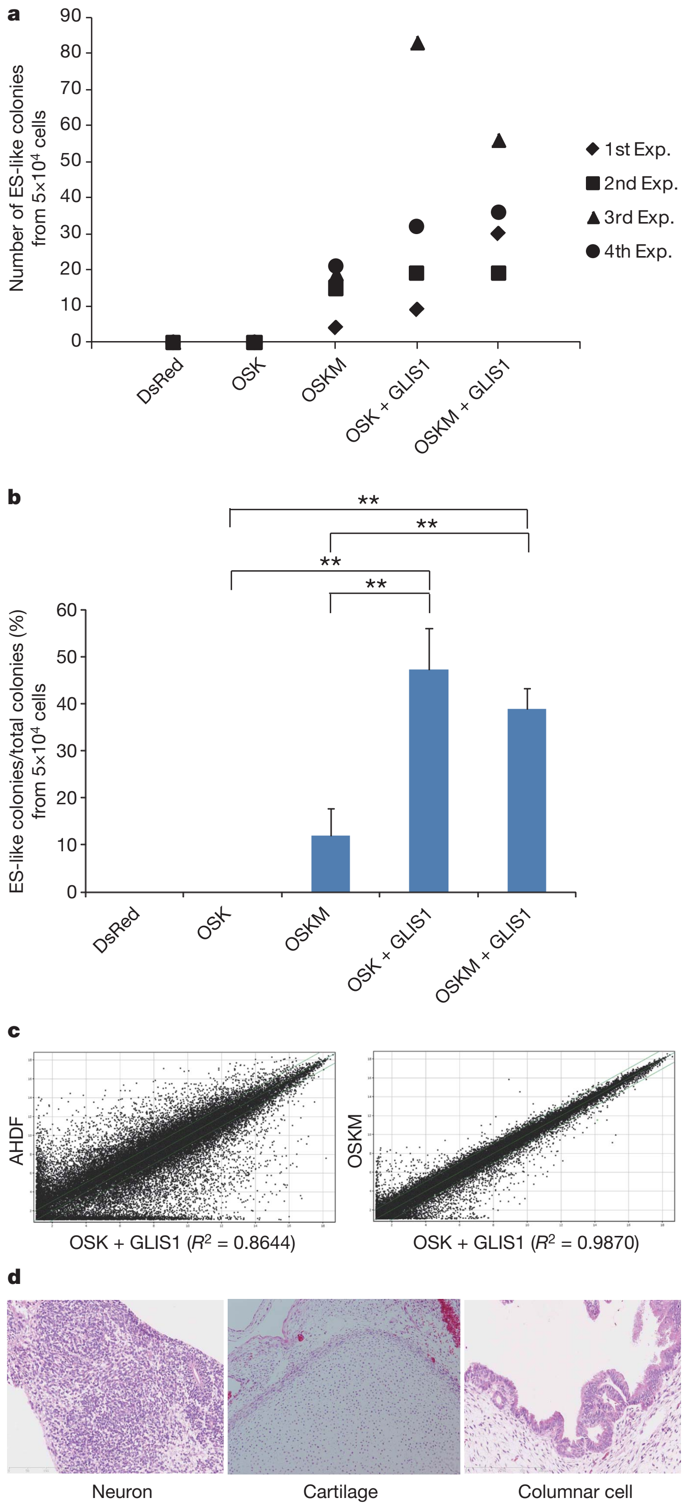


Figure 2 | Promotion of human iPSC generation by GLIS1. a, Number of ES-cell-like colonies from human dermal fibroblasts 30 d after infection. **b**, ES-cell-like colonies as a proportion of total colonies. **, $P < 0.01$ compared to cells expressing OSK alone. Error bars, s.d.; $n = 3$. **c**, Scatter plots comparing global gene expression patterns between iPSCs generated by OSK + GLIS1 and adult dermal fibroblasts (AHDF) (left panel), and between iPSCs from OSK + GLIS1 and iPSCs from OSKM (right panel). The green diagonal lines indicate twofold changes between the two samples. The correlation coefficient (R^2) is also shown. **d**, iPSCs generated by OSK + GLIS1 were subcutaneously transplanted into nude mice and teratomas were analysed histologically with haematoxylin and eosin staining.

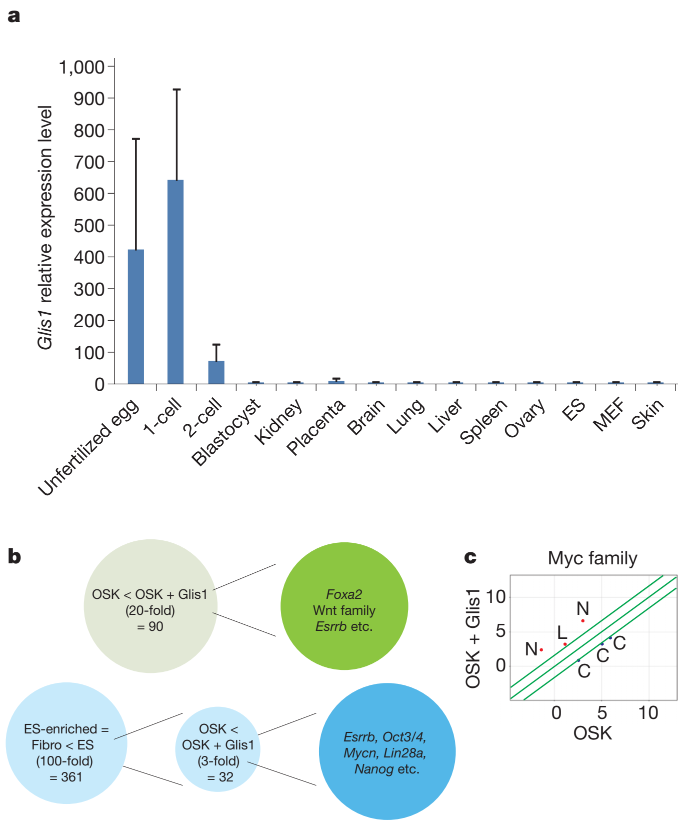


Figure 3 | Characterization of Glis1: expression and roles during iPSC generation. **a**, Expression patterns of *Glis1* in different mouse tissues. Data are normalized to glyceraldehyde-3-phosphate dehydrogenase expression; *Glis1* expression in the kidney is set at a relative level of 1. Error bars, s.d.; $n = 4$. **b**, Ninety genes were found to be upregulated more than 20-fold in OSK + Glis1 cells compared to OSK + mock cells (upper panel). These included *Foxa2*, multiple Wnt-family genes and *Esrrb*. We also focused on 361 probes for which expression was more than 100-fold higher in ES cells than in fibroblasts. Among these, 32 probes showed an expression level that was more than threefold higher in OSK + Glis1 cells than in OSK + mock cells (lower panel). These included *Esrrb, Oct3/4, Mycn, Lin28a* and *Nanog*. **c**, Expression levels of the Myc-family genes (C, *Myc*; N, *Mycn*; L, *Mycl1*) in OSK + Mock and OSK + Glis1 cells. The green diagonal lines indicate twofold changes between the two cell types.

reprogramming by inhibiting p53. If this is the case, Glis1 should not be able to promote iPSC generation in cells with a p53-null background. To test this hypothesis, we introduced OSK plus mock (control) or OSK plus Glis1 into either wild-type or p53-knockout MEFs, both containing the *Nanog*–GFP reporter. Five days after transduction, we measured the proportion of *Nanog*–GFP-positive cells by flow cytometry. We found that even in p53-knockout MEFs, in which the generation of *Nanog*–GFP-positive cells by OSK was increased about 10-fold (to about 2%), the addition of Glis1 further increased the proportion of GFP-positive cells up to about 17% (Supplementary Fig. 5). These data indicate that Glis1 promotes iPSC generation irrespective of p53.

We then used the very high reprogramming efficiency in cells with the p53-null background to elucidate the function of Glis1. We sorted and collected *Nanog*–GFP-positive cells 5 days after the transduction of OSK plus mock or OSK plus Glis1 into the p53-knockout MEFs. We then conducted microarray analysis to compare the gene expression

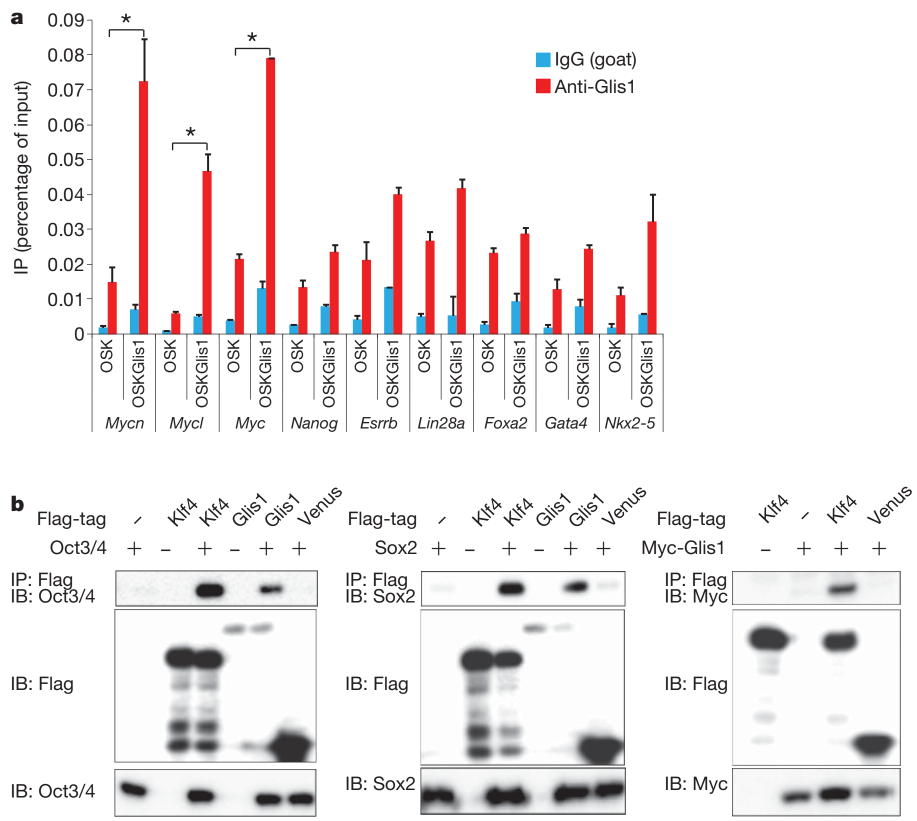


Figure 4 | Characterization of Glis1: target genes and protein–protein interactions. **a**, Chromatin immunoprecipitation and quantitative PCR analysis were conducted on the basis of microarray data, using a Glis1-specific antibody and PCR primers specific for *Mycn*, *Mycl1*, *Myc*, *Nanog*, *Esrrb*, *Lin28a* and *Foxa2*. GATA binding protein 4 (*Gata4*) and NK2 transcription factor related, locus 5 (*Nkx2-5*) were used as negative controls. IP, immunoprecipitate. Error bars, s.d.; $n = 2$; *, $P < 0.05$. **b**, Constructs encoding Flag-tagged Glis1 or Klf4 and untagged Oct3/4 (left panel), Flag-tagged Glis1 or Klf4 and untagged Sox2 (middle panel) or Flag-tagged Klf4 and Myc-tagged Glis1 (right panel) were transfected into HEK293T cells alone or in combination. Flag-tagged Venus was transfected as a negative control. The cell lysates were immunoprecipitated (IP) with an anti-Flag antibody, followed by an immunoblot analysis (IB). The expression levels in whole-cell lysates were determined by IB (bottom panels).

levels of these cell populations undergoing reprogramming (Fig. 3b, c and Supplementary Table 4). We found that Glis1 markedly increased the expression of several genes whose products have been shown to enhance iPSC generation. These included oestrogen-related receptor, beta (*Esrrb*)[13], several Wnt ligands (*Wnt3*, *Wnt6*, *Wnt8a* and *Wnt10a*)[14], lin-28 homologue A (*Lin28a*)[15], Nanog (ref. 16), *Mycn* and *Mycl1*(ref. 17). In contrast, the expression of *Myc* was suppressed by Glis1 (Fig. 3c). We have previously shown that Mycn and Mycl1 predominantly increase the numbers of ES-cell-like colonies, whereas Myc increases both ES-cell-like and non-ES-cell-like colonies[17]. Therefore, the altered balance between Mycn/Mycl1 and Myc should contribute, at least in part, to the specific promotion of iPSC generation by Glis1. Glis1 also markedly enhanced the expression of forkhead box A2 (*Foxa2*), a transcription factor that antagonizes the epithelial-to-mesenchymal transition. Because this transition is a prerequisite for iPSC generation[18,19], the activation of *Foxa2* should also have a role in the promotion of iPSC generation by Glis1. We confirmed the effect of Glis1 on *Nanog*, *Mycn*, *Myc*, neurogranin and tetraspanin 18 in a p53 wild-type background by quantitative PCR (Supplementary Fig. 6). Taken together, these data demonstrate that Glis1 promotes iPSC generation by activating multiple pro-reprogramming pathways.

We next performed chromatin immunoprecipitation assays to identify the direct transcriptional targets of Glis1. Cell lysates were isolated from p53-knockout MEFs transduced with OSK plus mock or OSK plus Glis1. Candidate target genes identified from the microarray analyses were amplified by PCR (Fig. 4a). We found that significantly higher amounts of *Mycn*, *Mycl1* and *Myc* were precipitated from the cells transduced with OSK plus Glis1 than from those transduced with OSK plus mock. In contrast, no such specific precipitation was observed with *Esrrb*, *Lin28a*, *Foxa2* or *Nanog*. These results indicate that the three *Myc* genes are direct targets of Glis1, whereas *Esrrb*, *Lin28a*, *Foxa2* and *Nanog* may be indirect targets.

We next examined whether Glis1 physically associates with the OSK proteins. Using Flag-tagged Glis1, we saw that Oct3/4 and Sox2 co-purified with Glis1 (Fig. 4b), whereas co-purification was not observed with a Flag-tagged Venus protein. In addition, we observed the co-purification of Flag–Klf4 with Myc-tagged Glis1 (Fig. 4b). The zinc-finger domain of Glis1 and its N-terminal region were required for the interaction with Klf4 (Supplementary Fig. 7). The interaction between Klf4 and Glis1 was further confirmed with an *in vitro* protein fragment complementation assay (Supplementary Fig. 8). These data indicate that Glis1 can associate with OSK by a protein–protein interaction and thereby might promote the activation of OSK target genes.

In contrast to oocytes and one-cell-stage embryos, we found that the expression of Glis1 was very low in ES cells. We therefore examined the effects of forced expression of Glis1 in mouse ES cells[20] and found that this suppressed their proliferation (Supplementary Fig. 9). This effect may have contributed to the smaller number of partially reprogrammed cells observed with OSK plus Glis1, because such cells would fail to silence retroviruses and would still express Glis1 transgenes, which would suppress proliferation.

This study shows that the transcription factor Glis1, which is highly enriched in unfertilized eggs and one-cell-stage embryos, promotes iPSC generation effectively and specifically by activating multiple pro-reprogramming pathways. Glis1 might thus be a link between reprogramming during iPSC generation and reprogramming after nuclear transfer. Furthermore, iPSCs generated by OSK and Glis1 did not cause a marked increase in mortality of chimaeric mice, although this did occur with iPSCs generated by Oct3/4, Sox2, Glis1 and Myc (Supplementary Fig. 10) and with iPSCs generated by OSK and Myc, as reported previously[17]. The identification of Glis1 might therefore be beneficial for future applications of iPSC technology.

METHODS SUMMARY

To screen transcription factors for their effects on iPSC generation, cDNAs were used from the human proteome expression resource (HuPEX) library[21]. Gateway entry clones of 1,437 human transcription factors were transferred to pMXs-GW retroviral expression vectors using the Gateway LR reaction. MEFs were isolated from 13.5 days post coitum (d.p.c.) embryos and adult skin fibroblasts were isolated from 20-week-old mice. The generation of mouse iPSCs with retroviruses was performed as described previously[2,9]. Human iPSCs were also generated as described previously[22]. The shRNA-mediated knockdown was performed as described in ref. 12. Retroviruses (pMXs) were generated with Plat-E packaging cells[23]. ES cells and iPSCs were cultured on SNL feeder cells[24]. The analyses of iPSCs, such as RT–PCR, alkaline phosphatase staining, DNA microarrays, *in vitro* differentiation, teratoma formation, bisulphite genomic sequencing and chimaera experiments, were performed as previously described[1,9,22]. Animal experiments were approved by committees of Kyoto University and the Japan Science and Technology Agency. To examine whether Glis1 is physically associated with the OSK proteins, immunoprecipitation and immunoblotting analyses were performed, as well as an *in vitro* protein fragment complementation assay[25]. In addition, a ChIP analysis was performed on Glis1 to identify its target genes. Sequences of primers and shRNAs are listed in Supplementary Tables 5 and 6, respectively. Microarray data are available through GEO with accession number GSE26431.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 May 2010; accepted 11 April 2011.

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
2. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotechnol.* **26,** 101–106 (2008).
3. Yamanaka, S. A fresh look at iPS cells. *Cell* **137,** 13–17 (2009).
4. Yamanaka, S. Strategies and new developments in the generation of patient-specific pluripotent stem cells. *Cell Stem Cell* **1,** 39–49 (2007).
5. Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460,** 49–52 (2009).
6. Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465,** 704–712 (2010).
7. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385,** 810–813 (1997).
8. Egli, D., Rosains, J., Birkhoff, G. & Eggan, K. Developmental reprogramming after chromosome transfer into mitotic mouse zygotes. *Nature* **447,** 679–685 (2007).
9. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germ-line competent induced pluripotent stem cells. *Nature* **448,** 313–317 (2007).
10. Kim, Y. S. *et al.* Identification of Glis1, a novel Gli-related, Kruppel-like zinc finger protein containing transactivation and repressor functions. *J. Biol. Chem.* **277,** 30901–30913 (2002).
11. Guo, G. *et al.* Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* **18,** 675–685 (2010).
12. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* **460,** 1132–1135 (2009).
13. Feng, B. *et al.* Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nature Cell Biol.* **11,** 197–203 (2009).
14. Marson, A. *et al.* Wnt signaling promotes reprogramming of somatic cells to pluripotency. *Cell Stem Cell* **3,** 132–135 (2008).
15. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318,** 1917–1920 (2007).
16. Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138,** 722–737 (2009).
17. Nakagawa, M., Takizawa, N., Narita, M., Ichisaka, T. & Yamanaka, S. Promotion of direct reprogramming by transformation-deficient Myc. *Proc. Natl Acad. Sci. USA* **107,** 14152–14157 (2010).
18. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).
19. Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7,** 51–63 (2010).
20. Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* **12,** 2048–2060 (1998).
21. Goshima, N. *et al.* Human protein factory for converting the transcriptome into an *in vitro*-expressed proteome. *Nature Methods* **5,** 1011–1017 (2008).
22. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
23. Morita, S., Kojima, T. & Kitamura, T. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.* **7,** 1063–1066 (2000).
24. McMahon, A. P. & Bradley, A. The *Wnt*-1 (*int*-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell* **62,** 1073–1085 (1990).
25. Hashimoto, J. *et al.* Novel *in vitro* protein fragment complementation assay applicable to high-throughput screening in a 1536-well format. *J. Biomol. Screen.* **14,** 970–979 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Yamamoto, Y. Yamada and the members of our laboratory for valuable scientific discussions and administrative support. We thank M. Nakagawa, H. Seki, M. Murakami, A. Okada, M. Narita, M. Inoue, H. Shiga and T. Matsumoto for technical assistance and H. Suemori (Kyoto University) for human ES cells. This work was supported in part by grants from the New Energy and Industrial Technology Development Organization (NEDO), the Leading Project of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program) of the Japanese Society for the Promotion of Science (JSPS), Grants-in-Aid for Scientific Research from JSPS and MEXT, and the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NIBIO). S.Y. is a member of scientific advisory boards of iPearian Inc. and iPS Academia Japan.

Author Contributions M.M. conducted most of the experiments in this study. K.Y. analysed the interactions of proteins. T.N. performed the computer analyses of the DNA microarray data, teratoma experiments, overexpression in ES cells and statistical analysis. R.S. generated mouse iPSCs and characterized mouse and human iPSCs. I.K. generated human iPSCs. T.I. performed the chimaera and teratoma experiments and maintained the mouse lines. Y.K. selected cDNA clones from HuPEX with bioinformatics. H.M. produced the retroviral expression clones. N.G. and S.Y. supervised the project. M.M. and S.Y. wrote the manuscript.

Author Information The microarray data are available from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) with the accession number GSE26431. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.Y. (yamanaka@cira.kyoto-u.ac.jp) or N.G. (n-goshima@aist.go.jp).

METHODS

cDNA library. cDNAs used to screen for novel factors that alter the efficacy of iPSC generation were obtained from the human proteome expression resource (HuPEX) library[21]. Among the 33,275 cDNAs, we selected those known to be transcription factors or identified by keyword searches of the Human Gene and Protein Database (HGPD, http://www.HGPD.jp/) and Entrez gene (http://www.ncbi.nlm.nih.gov/gene). We used cDNAs that covered more than 80% of the open reading frame reported in RefSeq and had identity with the reported protein sequence of more than 95% at the amino acid level. cDNAs encoding OCT3/4, SOX2, KLF4 or MYC were excluded. This resulted in 1,437 cDNAs (Supplementary Table 1), which were transferred to the pMXs-GW retroviral expression vector using the Gateway LR reaction.

Cell culture. Mouse iPSCs were maintained in ES cell medium (DMEM containing 15% fetal calf serum (FCS), $1\times$ non-essential amino acids (NEAA), 1 mM sodium pyruvate, 5.5 mM 2-mercaptoethanol (2-ME), 50 units ml^{-1} penicillin and 50 μg ml^{-1} streptomycin) on feeder layers of mitomycin-C-treated SNL cells stably expressing the puromycin-resistance gene[24]. As a source of leukaemia-inhibitory factor (LIF), we used the conditioned medium from Plat-E cell cultures that had been transduced with a LIF-expressing vector. Human iPSCs were generated and maintained in primate ES cell medium (ReproCELL), supplemented with 4 ng ml^{-1} recombinant human basic fibroblast growth factor, 50 units ml^{-1} penicillin and 50 μg ml^{-1} streptomycin. MEFs, mouse skin fibroblasts and human fibroblasts were maintained in DMEM containing 10% FCS, 50 units ml^{-1} penicillin and 50 μg ml^{-1} streptomycin. Plat-E cells[23] were maintained in DMEM containing 10% FCS, 50 units ml^{-1} penicillin, 50 μg ml^{-1} streptomycin, 1 μg ml^{-1} puromycin and 10 μg ml^{-1} blasticidin S. We used 13.5 d.p.c. embryos for MEF isolation and 20-week-old mice for the isolation of skin fibroblasts.

Mouse iPSC generation. The generation of mouse iPSCs with retroviruses was performed as previously described[2,9] with some modifications. Briefly, Plat-E cells were seeded at 2.5×10^6 cells per 100-mm dish. On the next day, pMXs-based retroviral vectors for each gene were independently introduced into Plat-E cells using the FuGENE 6 transfection reagent. After 24 h, the medium was replaced with 10 ml of DMEM containing 10% FCS. Fibroblasts were seeded at 8×10^5 cells per dish, in 100-mm dishes covered with a layer of gelatin or feeder cells. The next day, virus-containing supernatants from the Plat-E cultures were recovered and mixed together (for example, viruses for *Oct3/4*, *Sox2*, *Klf4* and *Glis1*). Fibroblasts were incubated in the virus/polybrene-containing supernatants at a final concentration of 4 μg ml^{-1} for 24 h. Three days after infection, the medium was changed to ES cell medium supplemented with LIF. Fibroblasts on gelatin-coated dishes were then re-seeded onto dishes with feeder cells. The shRNA-mediated knockdown was performed as previously described[12].

Generation of human iPSCs. Human iPSCs were generated as previously described[22] with some modifications. Briefly, Plat-E cells were plated at 3.6×10^6 cells per 100-mm dish. The next day, pMXs-based retroviral vectors for each gene were independently introduced into the Plat-E cells using the FuGENE 6 transfection reagent. After 24 h, the medium was replaced with new medium. Human fibroblasts expressing the mouse *Slc7a1* (solute carrier family 7 (cationic amino acid transporter, y+ system), member 1) gene were seeded at 8×10^5 cells per 100-mm dish. The next day, virus-containing supernatants were recovered and mixed together. Fibroblasts were incubated in the virus/polybrene-containing supernatants at a final concentration of 4 μg ml^{-1} for 24 h. Six days after transduction, fibroblasts were harvested by trypsinization and replated at 5×10^4 or 5×10^5 cells per 100-mm dish on SNL feeder cells. The next day, the medium was replaced with primate ES cell medium supplemented with 4 ng ml^{-1} basic fibroblast growth factor.

Characterization of iPSCs. The RT–PCR analyses, alkaline phosphatase staining, *in vitro* differentiation, teratoma formation, bisulphite genomic sequencing and chimaera experiments were performed as previously described[1,9,22]. The primers used for RT–PCR are listed in Supplementary Table 5. In the *in vitro* differentiation assay, differentiated cells were stained positive for α-fetoprotein (endoderm), α-smooth muscle actin (mesoderm) and nestin (ectoderm). Nuclei were stained with Hoechst. For bisulphite genomic sequencing, the white circles indicate unmethylated CpG dinucleotides, whereas the black circles indicate methylated CpG dinucleotides.

DNA microarray. Total RNAs were labelled with Cy3 and hybridized to either a Whole Mouse Genome Microarray or a Whole Human Genome Microarray (Agilent) according to the manufacturer's protocol. Arrays were scanned using the G2505C Microarray Scanner System (Agilent). The data were analysed using the GeneSpring GX11.0.1 software program (Agilent). The microarray data are available from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) with the accession number GSE26431.

Chromatin immunoprecipitation assay. We used the Active Motif ChIP-IT Express kit for the chromatin immunoprecipitation assay. Genomic DNA and nuclear proteins were fixed with formaldehyde. Immunoprecipitation was performed with either anti-Glis1 (Santa Cruz) or purified goat IgG antibody and the elutes were used as templates for quantitative PCR. We selected DNA fragments containing putative Glis1-binding sites for PCR amplification. The primers used for quantitative PCR in the ChIP assay are listed in Supplementary Table 5.

Immunoprecipitation and immunoblotting analyses. Because the expression levels of *Glis1* in ES cells and fibroblasts are low, we were not able to elucidate whether there was an association among the endogenous proteins. HEK293T cells were therefore transfected with each cDNA clone in an expression vector and were lysed in CytoBuster (Novagen). Cell lysates were incubated with an anti-Flag M2 Affinity Gel (Sigma) for 2 h and then removed. The gel suspensions were boiled in sample buffer and analysed by SDS–polyacrylamide gel electrophoresis and immunoblotting. The immunoblot analyses were performed using the following antibodies: anti-Flag M2 (Sigma), anti-Myc (Roche), anti-Oct3/4 (Santa Cruz) and anti-Sox2 (MBL).

***In vitro* protein fragment complementation assay.** We prepared split monomeric Kusabira-Green protein (mKG) fragment proteins (Amalgaam) fused to Glis1 and Klf4 using a wheat-germ cell-free protein synthesis system (CellFree Sciences)[25]. Each protein solution was dispensed into a 384-well plate. After incubation at 25 °C for 8 h or 23 h, the fluorescence was measured using the Typhoon 9200 (GE Healthcare).

Overexpression of genes in ES cells. The mouse ES cell line MG1.19 was maintained in DMEM containing 10% FCS, $1\times$ NEAA, 1 mM sodium pyruvate, 5.5 mM 2-ME, 50 units ml^{-1} penicillin, 50 μg ml^{-1} streptomycin and LIF. The vectors pCAG-IP (Mock) or pCAG-Glis1-IP were introduced into MG1.19 cells using Lipofectamine 2000 on day -1. On day zero, 1×10^5 cells were re-seeded on a gelatin-coated 6-well plate. On day 4, the cell number was counted.

Statistical analyses. A one-way repeated-measures ANOVA and a post-hoc Bonferroni test were used for the analyses of the data in Figs 1c and 2b. The unpaired *t*-test was used for statistical analysis of the data shown in Fig. 4a (between OSK and OSKGlis1). Differences were considered to be statistically significant for *P*-values <0.05 (*), <0.01 (**) or <0.001 (***).

doi:10.1038/nature10155

A function for cyclin D1 in DNA repair uncovered by protein interactome analyses in human cancers

Siwanon Jirawatnotai[1,2], Yiduo Hu[1]*, Wojciech Michowski[1]*, Joshua E. Elias[3]†, Lisa Becks[1,4], Frederic Bienvenu[1]†, Agnieszka Zagozdzon[1], Tapasree Goswami[3], Yaoyu E. Wang[5], Alan B. Clark[6], Thomas A. Kunkel[6], Tanja van Harn[1], Bing Xia[7], Mick Correll[5], John Quackenbush[5,8], David M. Livingston[1], Steven P. Gygi[3] & Piotr Sicinski[1]

Cyclin D1 is a component of the core cell cycle machinery[1]. Abnormally high levels of cyclin D1 are detected in many human cancer types[2]. To elucidate the molecular functions of cyclin D1 in human cancers, we performed a proteomic screen for cyclin D1 protein partners in several types of human tumours. Analyses of cyclin D1 interactors revealed a network of DNA repair proteins, including RAD51, a recombinase that drives the homologous recombination process[3]. We found that cyclin D1 directly binds RAD51, and that cyclin D1–RAD51 interaction is induced by radiation. Like RAD51, cyclin D1 is recruited to DNA damage sites in a BRCA2-dependent fashion. Reduction of cyclin D1 levels in human cancer cells impaired recruitment of RAD51 to damaged DNA, impeded the homologous recombination-mediated DNA repair, and increased sensitivity of cells to radiation *in vitro* and *in vivo*. This effect was seen in cancer cells lacking the retinoblastoma protein, which do not require D-cyclins for proliferation[4,5]. These findings reveal an unexpected function of a core cell cycle protein in DNA repair and suggest that targeting cyclin D1 may be beneficial also in retinoblastoma-negative cancers which are currently thought to be unaffected by cyclin D1 inhibition.

To elucidate the molecular functions of cyclin D1 in human cancer cells, we set out to identify cyclin D1-interacting proteins in four types of human tumours: mantle cell lymphoma (Granta 519 cells), breast cancer (MCF7 and ZR-75-1), squamous cell carcinoma (UMSCC-2), and colorectal cancer (HT-29). These cell lines overexpress cyclin D1 due to amplification or rearrangements within the cyclin D1 gene (*CCND1*), or due to mutations within the cyclin D1 degradation machinery[6–9]. Cyclin D1-containing complexes were purified using double immunoaffinity purification[10] (Supplementary Fig. 1), and the identity of cyclin D1 interactors was determined by repeated rounds of liquid chromatography and high-throughput mass spectrometry (LC-MS/MS). A total of 132 proteins were identified with high confidence (Fig. 1a, Supplementary Fig. 2 and Supplementary Tables 1–5).

We constructed a biological process/molecular function enrichment heat map of cyclin D1 interactors (Fig. 1b and Supplementary Table 6), and searched for functions enriched in all four tumour types. Unexpectedly, apart from cell cycle control proteins, we observed the DNA repair category amongst the most enriched functions (Fig. 1b, Supplementary Fig. 3a and Supplementary Table 6). Using published databases of physical and functional interactions, we constructed a protein interaction network of all cyclin D1 interactors encountered by us in the five cell lines, and observed a cluster of DNA repair proteins centred on RAD51, a key DNA recombinase mediating DNA repair via homologous recombination[3] (Supplementary Fig. 3b). These observations suggested that cyclin D1 may play a role in DNA damage repair.

Consistent with this possibility, we found that depletion of cyclin D1 by short hairpin RNA (shRNA) in cervical carcinoma HeLa cells and lung cancer H2009 cells, significantly increased the sensitivity of cancer cells to ionizing radiation (Fig. 2a and Supplementary Fig. 4a, c). We used these two retinoblastoma protein (pRB)-negative cell lines to rule out cell cycle effects of cyclin D1 knockdown, as pRB-negative cancer cells do not require D-cyclins for proliferation[4,5] (Supplementary Fig. 5). Reduction of cyclin D1 levels also sensitized cancer cells to DNA-damaging drugs camptothecin and etoposide (Fig. 2b and Supplementary Fig. 4d). Moreover, mouse embryonic fibroblasts lacking D-cyclins ($D1^{-/-}D2^{-/-}D3^{-/-}$)[11], which have normal cell cycle progression, showed increased sensitivity to ionizing radiation; re-expression of cyclin D1 in these cells restored radiation sensitivity (Supplementary Figs 4e, f, 6).

Cyclin D1 activates cyclin-dependent kinases CDK4 and CDK6 (ref. 1). However, treatment of HeLa cells with PD 0332991, a specific inhibitor of cyclin D-CDK4/6 kinase[12], had no effect on sensitivity of cancer cells to ionizing radiation (Supplementary Fig. 4b). Moreover, expression of cyclin D1(K112E) point-mutant, which is unable to activate the CDKs[13,14], restored radiation susceptibility in cyclin D-null mouse embryonic fibroblasts (Supplementary Fig. 6). Collectively, these observations indicated that cyclin D1 might have a kinase-independent function in DNA repair.

We used comet assay[15] to directly assess the efficiency of DNA repair in cyclin D1-depleted human cancer cells. We found that radiation induced comparable levels of DNA damage in control and in cyclin D1-depleted cells (4-h time-point in Fig. 2c and in Supplementary Fig. 7a). However, at 16 h post-radiation, more unrepaired DNA persisted in cyclin D1-depleted cells (Fig. 2c and Supplementary Fig. 7). We verified that this impaired rate of DNA repair was not a result of grossly compromised DNA damage-induced signalling (Supplementary Fig. 8a, b).

Because among cyclin D1 interactors we observed a network of DNA repair proteins including RAD51, a key DNA recombinase that drives the homologous recombination process[3,16] (Supplementary Fig. 3b), we investigated a possible functional connection between cyclin D1 and homologous recombination-dependent DNA repair. Using a homologous recombination repair reporter, the DR-GFP system[17], we found that depletion of cyclin D1 in HeLa and H2009 cells significantly reduced the homologous recombination rate (Fig. 2d and

[1]Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02215, USA. [2]Institute of Molecular Biosciences, Mahidol University, Salaya, Nakhon Prathom, 73170 Thailand. [3]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Department of Pharmacology, Massachusetts College of Pharmacy and Health Sciences, Boston, Massachusetts 02115, USA. [5]Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [6]Laboratory of Molecular Genetics and Laboratory of Structural Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina 27709, USA. [7]Department of Radiation Oncology, The Cancer Institute of New Jersey, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, New Brunswick, New Jersey 08903, USA. [8]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, and Harvard School of Public Health, Boston, Massachusetts 02215, USA. †Present addresses: Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California 94305, USA (J.E.E.); Institute of Functional Genomics, UMR 5203 CNRS – U 661 INSERM – Université de Montpellier, 34094 Montpellier, France (F.B.).
*These authors contributed equally to this work.

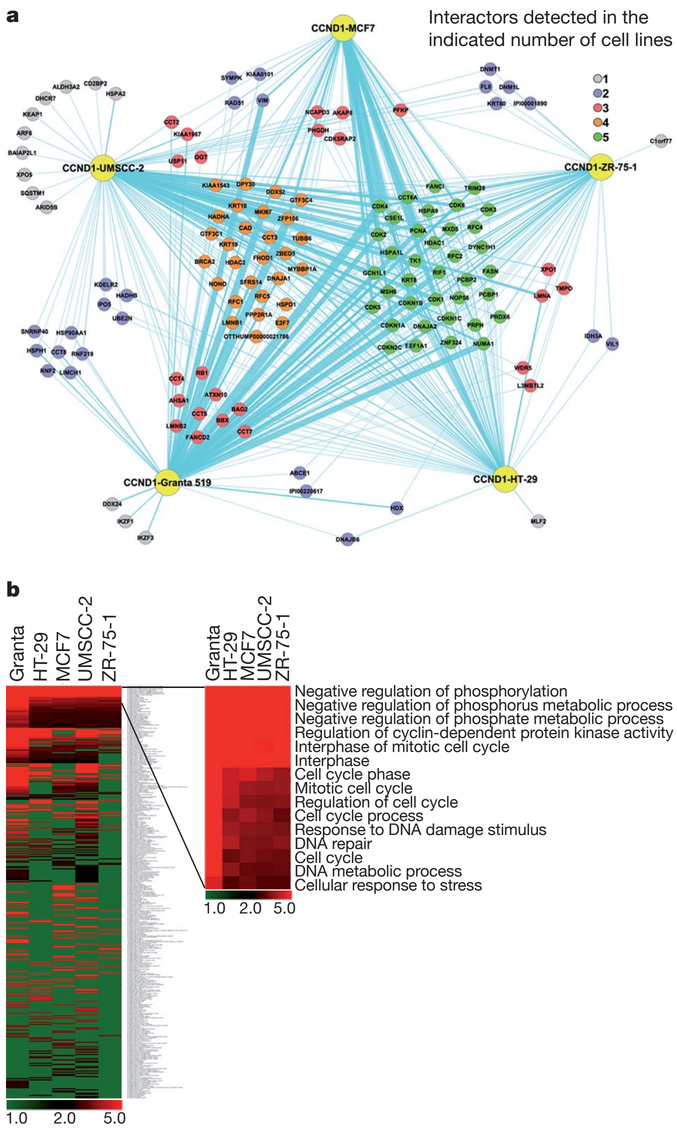


Figure 1 | Analyses of cyclin D1 interactors identified in human cancers. **a**, Diagram depicting cyclin D1 interactors, grouped by cell line in which they were detected. Line thickness corresponds to the abundance of peptides for each protein detected by mass spectrometry in a given cell line. **b**, Biological process/molecular function enrichment heat map of cyclin D1 interactors. Each column corresponds to the indicated cell line, rows denote distinct biological process/molecular functions. Red, enriched functions; green, functions not enriched. Left panel shows a complete map, right panel an enlarged map of molecular functions/biological processes enriched across all five cancer cell lines analysed.

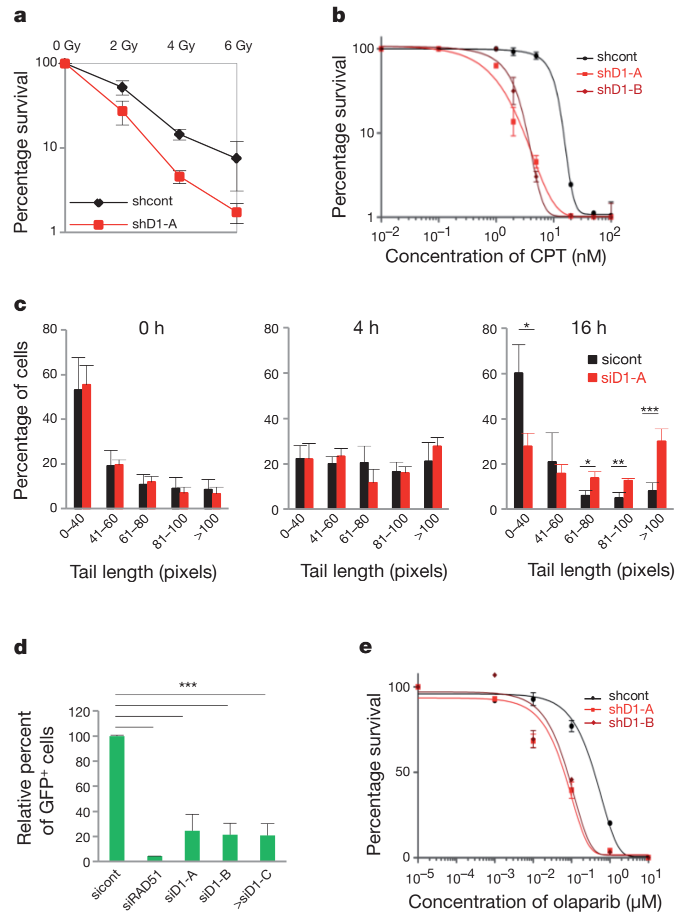


Figure 2 | Impaired DNA repair upon reduction of cyclin D1 levels. **a**, Colony survival assay of HeLa cells expressing anti-cyclin D1 shRNA (shD1-A) or control shRNA (shcont), after the indicated doses of ionizing radiation. IC_{50} (half maximal inhibitory concentration/dose) = 2.15 Gy (shcont), 1.07 Gy (shD1-A). **b**, **e**, Colony survival assays of H2009 cells expressing anti-cyclin D1 shRNAs (shD1-A or shD1-B), or control shRNA (shcont), after exposure to indicated doses of camptothecin (CPT) or AZD2281 (olaparib). IC_{50}s for camptothecin: 15.1 nM (shcont), 2.8 nM (shD1-A), 3.4 nM (shD1-B); olaparib: 0.39 μM (shcont), 0.083 μM (shD1-A), 0.065 μM (shD1-B). **c**, Results of comet assays to measure DNA damage repair in cyclin D1-depleted HeLa cells (siD1-A) and control cells (sicont). Shown are percentages of cells containing a given comet tail length at the indicated time-points after ionizing radiation. **d**, Homologous recombination assay in HeLa cells using DR-GFP reporter system. sicont, control siRNA; siRAD51, anti-RAD51 siRNA; siD1-A, -B,-C, anti-cyclin D1 siRNAs. $*P \leq 0.05$; $**P \leq 0.01$; $***P \leq 0.005$; error bars, standard deviation; $n = 3$–5.

Supplementary Fig. 9). As expected, re-expression of small interfering RNA (siRNA)-resistant cyclin D1 rescued this effect (Supplementary Fig. 10a–c). These results indicated that cyclin D1 is required for an efficient homologous recombination DNA repair process. Consistent with this notion, cyclin D1 depletion sensitized cancer cells to treatment with poly (ADP-ribose) polymerase (PARP) inhibitor (Fig. 2e), in concordance with the reports that deficiency in homologous recombination renders cells hypersensitive to these agents[18].

To investigate the function of cyclin D1 in homologous recombination, we focused on cyclin D1 interaction with RAD51, detected in our screen. We established that cyclin D1 binds directly to RAD51 using purified, recombinant proteins, and that the amino terminus of cyclin D1 and the carboxy terminus of RAD51 mediate the interaction (Fig. 3a and Supplementary Fig. 11). Expression of a cyclin D1 mutant which is deficient in RAD51 binding failed to restore normal homologous recombination rate in cyclin D1-depleted cells (Supplementary Fig. 10d). We also verified physical binding of the endogenous cyclin D1 and RAD51 proteins in several human cancer cell lines (Fig. 3b and Supplementary Fig. 12a). The interaction between cyclin D1 and RAD51 was induced by radiation and it intensified with the increased ionizing radiation dose (Fig. 3c, upper panel, and Supplementary Fig. 12b, c). As reported, radiation resulted in a decrease of total levels of cyclin D1, without major changes in RAD51 levels[19,20] (Fig. 3c, lower panel, and Supplementary Fig. 12c, lower panel). The strong induction of cyclin D1–RAD51 interaction after radiation, in the face of reduced total cyclin D1 levels, suggests that an increased number of remaining cyclin D1 molecules is recruited to RAD51 complex to facilitate the DNA repair process. In pRB-positive cancer cells, decrease of total cyclin D1 levels upon radiation contributes to cell cycle arrest of these cells[19]. Conversely, forced overexpression of cyclin D1 may overcome cell cycle checkpoints, and may lead to radio-resistant DNA synthesis and apoptosis[20,21]. Our results demonstrate that after radiation the remaining pool of cyclin D1 has a positive role in DNA repair.

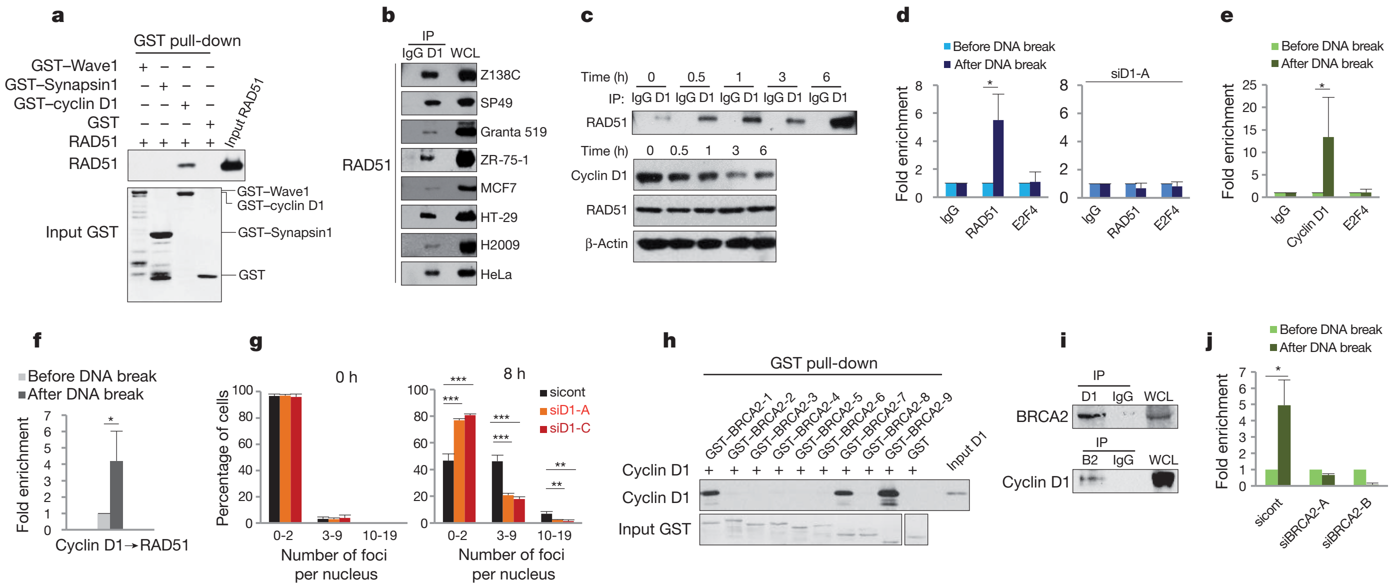


Figure 3 | Functional interaction between cyclin D1 and RAD51. a, Direct binding between GST–cyclin D1 and recombinant RAD51 protein. GST–Synapsin1, GST–Wave1 and GST served as negative controls for RAD51 binding. Upper panel: RAD51 was detected by immunoblotting. Lower panel: input GST–proteins visualized by Ponceau staining. **b**, Interaction between endogenous cyclin D1 and RAD51, detected by cyclin D1 immunoprecipitation and anti-RAD51 immunoblotting, in indicated cell lines. **c**, Cyclin D1–RAD51 interaction at different time-points after irradiating HeLa cells. Upper panel: cyclin D1 immunoprecipitation–RAD51 immunoblotting. Lower panel: levels of cyclin D1 and RAD51 after ionizing radiation, determined by western blotting. **d**, Left panel: recruitment of RAD51 to I-SceI-induced double-stranded DNA break, gauged by anti-RAD51 ChIP followed by PCR with primers adjacent to DNA damage site. Bars show enrichment around DNA damage site before and after induction of a DNA break. For control, we used anti-E2F4 and non-immune IgG (IgG) ChIP followed by PCR with the same set of primers adjacent to DNA damage site. Right panel: same assay in cells expressing anti-cyclin D1 siRNA (sD1-A). **e**, Recruitment of cyclin D1 to I-SceI-induced double-stranded DNA break, tested as in **d** using anti-cyclin D1 ChIP. Bars show enrichment around DNA damage site. **f**, Co-localization of cyclin D1 and RAD51 at the DNA damage site. Anti-cyclin D1 ChIP was followed by anti-RAD51 re-ChIP and PCR with primers adjacent to double-stranded DNA break. **g**, Reduction of RAD51 recruitment to DNA damage foci in HeLa cells depleted of cyclin D1 (siD1-A, -C). Shown is percentage of cells with a given number of RAD51 foci per nucleus before (0 h) and 8 h after irradiation. **h**, *In vitro* binding assays using the indicated GST–BRCA2 fragments[26] and recombinant cyclin D1. Cyclin D1 was detected by anti-cyclin D1 immunoblotting. Amounts of input GST–BRCA2 fragments were verified by Ponceau staining (lower panel). **i**, Interaction between endogenous cyclin D1 and BRCA2 (B2) in H2009 cells, detected by immunoprecipitation–immunoblotting. **j**, Recruitment of cyclin D1 to an I-SceI-induced double-stranded DNA break, tested as in **e**, using cells transfected with independent anti-BRCA2 siRNAs (siBRCA2-A or siBRCA2-B), or with control siRNA (sicont). Bars show enrichment around DNA damage site. IP, immunoprecipitation; WCL, whole cell lysate; *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.005$; error bars, standard deviation; $n = 3$–5.

To test whether cyclin D1 is recruited to the sites of damaged DNA, where RAD51 localizes after radiation, we used a system in which a unique I-SceI recognition site has been integrated into the genome of HeLa cells[17]. Using this system, one can monitor recruitment of repair proteins to the site of DNA damage using targeted chromatin immunoprecipitation (ChIP), followed by PCR with primers adjacent to DNA damage site[22–24]. As expected, we observed recruitment of RAD51 to the DNA damage site (Fig. 3d, left panel and Supplementary Fig. 13a, left panel, d). Importantly, targeted anti-cyclin D1 ChIP showed that cyclin D1 was also recruited to the site of DNA damage (Fig. 3e and Supplementary Fig. 13f); this recruitment disappeared once cyclin D1 had been knocked down (Supplementary Fig. 13b, g). Moreover, ChIP of cyclin D1 followed by re-ChIP with anti-RAD51 antibody indicated that cyclin D1 and RAD51 colocalize on the sites of double-stranded DNA breaks (Fig. 3f and Supplementary Fig. 13c). Cyclin D1 recruitment to the damaged DNA was also detected by co-immunofluorescent staining with γ-H2AX at laser-induced DNA damage stripes (Supplementary Fig. 14).

To investigate how cyclin D1 facilitates the homologous recombination process, we studied recruitment of RAD51 to damaged DNA, a step that is required for RAD51 to perform its DNA recombinase function[25]. We found that knockdown of cyclin D1 reduced recruitment of RAD51 to I-SceI-induced DNA breaks (Fig. 3d, right panel and Supplementary Fig. 13a, right panel, e), to laser-induced DNA-damage stripes (Supplementary Fig. 15), and to radiation-induced DNA damage foci (Fig. 3g and Supplementary Figs 16 and 17). We verified that the overall levels of RAD51 and of several other DNA damage response proteins were not affected by cyclin D1 knockdown (Supplementary Fig. 8c). Collectively, these findings suggest that cyclin D1 helps to recruit RAD51 to DNA damage sites through a direct cyclin D1–RAD51 physical interaction.

We next asked how cyclin D1 is recruited to sites of DNA damage. Among cyclin D1 interactors we observed BRCA2 (Fig. 1a, Supplementary Fig. 3 and Supplementary Tables 1–5), an essential homologous recombination protein which is recruited to DNA damage sites prior to RAD51 (ref. 25). Using purified recombinant cyclin D1 protein and BRCA2 protein fragments[26], we established that cyclin D1 interacts directly with BRCA2 (Fig. 3h), and we verified physical interaction of the endogenous proteins (Fig. 3i). Importantly, we found that knockdown of BRCA2 reduced recruitment of cyclin D1 to DNA damage sites (Fig. 3j, Supplementary Fig. 18a). Depletion of cyclin D1 had no observable effect on BRCA2 recruitment (Supplementary Fig. 19a), consistent with the notion that BRCA2 acts upstream of cyclin D1. Likewise, depletion of cyclin D1 had no detectable impact on other upstream events, such as DNA end-resection (Supplementary Fig. 20a–d) and loading of a single-stranded DNA binding protein RPA34 (Supplementary Fig. 20e, f). In contrast, depletion of cyclin D1 inhibited downstream events, namely recruitment of RAD51, and led to decreased co-localization of RAD51 and BRCA2 at DNA damage sites (Supplementary Fig. 18b).

Collectively, these results are consistent with the model that cyclin D1 is recruited to DNA damage sites through BRCA2; cyclin D1 then helps to recruit RAD51 through a direct cyclin D1–RAD51 interaction. Depletion of cyclin D1 reduces RAD51 recruitment and reduces RAD51–BRCA2 co-localization at DNA damage sites, leading to impaired homologous recombination rate. Of note, a recent report

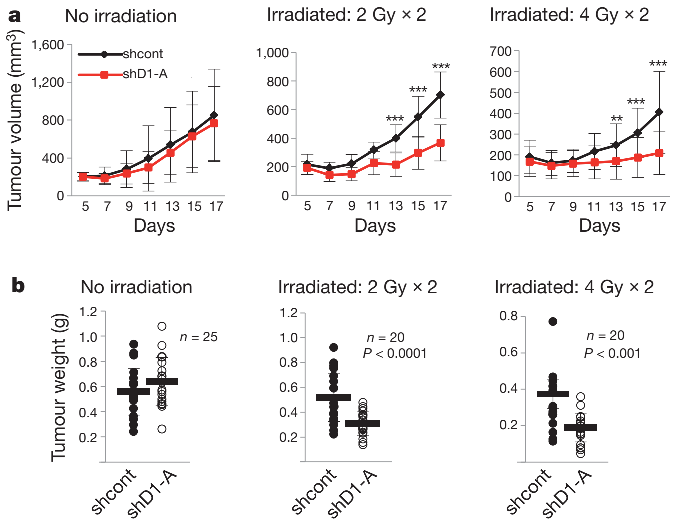


Figure 4 | Increased radiation-sensitivity of tumours with reduced cyclin D1 levels. **a**, H2009 cells expressing control shRNA (shcont) or anti-cyclin D1 shRNA (shD1-A) were injected into nude mice and tumour growth was monitored on the indicated days. Left panel, mice received no radiation. Middle and right panels: at days 5 and 10, tumours were exposed to 2 Gy or 4 Gy of ionizing radiation. Error bars, standard deviation; $^{**}P \leq 0.01$; $^{***}P \leq 0.005$. **b**, Weight of tumours at the end of observation period. Each circle represents an individual tumour, horizontal bars show mean values. *n*, number of mice analysed.

indicated that cyclin D1 has a role in DNA damage response and that forced targeting of cyclin D1 to chromatin resulted in increased co-recruitment of RAD51 upon DNA damage[27].

We verified that cyclin D1 is not required for recruitment of other DNA damage proteins detected as cyclin D1 interactors in our screen such as FANCD2, FANCI, PCNA and MSH6, as well as BRCA1 and MRE11 (Supplementary Figs 19, 21). We also found that depletion of cyclin D1 has no detectable impact on DNA mismatch repair rate (Supplementary Fig. 22).

Our observations suggested that depletion of cyclin D1 could sensitize human cancers to radiation, by limiting DNA repair. To test this, we knocked down cyclin D1 in three pRB-negative cancer cell lines: H2009, HeLa and in prostate cancer DU145. Cells were then injected subcutaneously into nude mice, which resulted in formation of tumours. As expected, depletion of cyclin D1 did not affect proliferation of cancer cells *in vitro* (Supplementary Figs 5c, d, 23a–c and 24a–c), and had no impact on the rate of tumour growth *in vivo* (Fig. 4a, b left panels and Supplementary Figs 24d–g left panels, and Supplementary Fig. 25a). In contrast, upon irradiation, cyclin D1-depleted tumours had retarded growth compared to control tumours, revealing increased sensitivity to ionizing radiation (Fig. 4a, b, middle and right panels, and Supplementary Figs 23d, e, 24d–g middle and right panels, 25b–e and 26–28, and data not shown). Hence, whereas cyclin D1 is dispensable for proliferation of pRB-negative tumour cells, it has an important role in DNA repair, which becomes evident once cells undergo DNA damage.

The cyclin D1 gene corresponds to the second most frequently amplified region within the human cancer genome[28]. Our proteomic screen revealed that cyclin D1 has a function in human cancer cells in DNA repair through homologous recombination. Thus, targeting cyclin D1 in combination with radiation treatment may have potential therapeutic value in a large pool of pRB-negative cancers.

METHODS SUMMARY

Affinity immunopurification and mass spectrometry were as described[29]. Comet assays were performed according to manufacturer's protocol (Trevigen). RAD51 and cyclin D1 ChIP were performed using anti-RAD51 (H-92) or anti-cyclin D1 antibody (H-295, Santa Cruz). For ChIP-re-ChIP, haemagglutinin-tagged cyclin D1 was ChIP with anti-haemagglutinin antibody (12CA5, Covance), eluted with haemagglutinin peptide, and re-ChIP with anti-RAD51 antibody. Glutathione-*S*-transferase-tagged proteins were based on pGEX-5x-3 (GE Healthcare). Cancer cells used for *in vivo* tumour study were transduced with lentiviruses expressing anti-cyclin D1 or control shRNAs. Tumour cells (10^7) were injected subcutaneously into *nu/nu* mice. Tumours were target-irradiated using ^{137}Cs as a source. Tumour size was assessed on indicated days, and tumours were weighed at the end of the experiments. Statistical significance of the differences was evaluated using paired two-tailed Student's *t*-test.

Received 13 July 2010; accepted 21 April 2011.

1. Sherr, C. J. & Roberts, J. M. Living with or without cyclins and cyclin-dependent kinases. *Genes Dev.* **18,** 2699–2711 (2004).
2. Deshpande, A., Sicinski, P. & Hinds, P. W. Cyclins and cdks in development and cancer: a perspective. *Oncogene* **24,** 2909–2915 (2005).
3. Baumann, P. & West, S. C. Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. *Trends Biochem. Sci.* **23,** 247–251 (1998).
4. Bates, S. *et al.* Absence of cyclin D/cdk complexes in cells lacking functional retinoblastoma protein. *Oncogene* **9,** 1633–1640 (1994).
5. Lukas, J. *et al.* Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. *Nature* **375,** 503–506 (1995).
6. Ek, S., Ortega, E. & Borrebaeck, C. A. K. Transcriptional profiling and assessment of cell lines as *in vitro* models for mantle cell lymphoma. *Leuk. Res.* **29,** 205–213 (2005).
7. Lin, D. I. *et al.* Phosphorylation-dependent ubiquitination of cyclin D1 by the SCF$^{\text{EBX4-}\alpha\text{B crystallin}}$ complex. *Mol. Cell* **24,** 355–366 (2006).
8. Hosokawa, Y. & Arnold, A. Mechanism of *cyclin D1* (*CCND1*, *PRAD1*) overexpression in human cancer cells: analysis of allele-specific expression. *Genes Chromosom. Cancer* **22,** 66–71 (1998).
9. Bartkova, J. *et al.* Abnormal patterns of D-type cyclin expression and G1 regulation in human head and neck cancer. *Cancer Res.* **55,** 949–956 (1995).
10. Nakatani, Y. & Ogryzko, V. Immunoaffinity purification of mammalian protein complexes. *Methods Enzymol.* **370,** 430–444 (2003).
11. Kozar, K. *et al.* Mouse development and cell proliferation in the absence of D-cyclins. *Cell* **118,** 477–491 (2004).
12. Toogood, P. L. *et al.* Discovery of a potent and selective inhibitor of cyclin-dependent kinase 4/6. *J. Med. Chem.* **48,** 2388–2406 (2005).
13. Hinds, P. W., Dowdy, S. F., Eaton, E. N., Arnold, A. & Weinberg, R. A. Function of a human cyclin gene as an oncogene. *Proc. Natl Acad. Sci. USA* **91,** 709–713 (1994).
14. Landis, M. W., Pawlyk, B. S., Li, T., Sicinski, P. & Hinds, P. W. Cyclin D1-dependent kinase activity in murine development and mammary tumorigenesis. *Cancer Cell* **9,** 13–22 (2006).
15. Ostling, O. & Johanson, K. J. Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells. *Biochem. Biophys. Res. Commun.* **123,** 291–298 (1984).
16. Baumann, P., Benson, F. E. & West, S. C. Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions *in vitro*. *Cell* **87,** 757–766 (1996).
17. Pierce, A. J., Johnson, R. D., Thompson, L. H. & Jasin, M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.* **13,** 2633–2638 (1999).
18. Lord, C. J. & Ashworth, A. Targeted therapy for cancer using PARP inhibitors. *Curr. Opin. Pharmacol.* **8,** 363–369 (2008).
19. Agami, R. & Bernards, R. Distinct initiation and maintenance mechanisms cooperate to induce G1 cell cycle arrest in response to DNA damage. *Cell* **102,** 55–66 (2000).
20. Pontano, L. L. *et al.* Genotoxic stress-induced cyclin D1 phosphorylation and proteolysis are required for genomic stability. *Mol. Cell. Biol.* **28,** 7245–7258 (2008).
21. Coco Martin, J. M., Balkenende, A., Verschoor, T., Lallemand, F. & Michalides, R. Cyclin D1 overexpression enhances radiation-induced apoptosis and radiosensitivity in a breast tumor cell line. *Cancer Res.* **59,** 1134–1140 (1999).
22. Sugawara, N., Wang, X. & Haber, J. E. *In vivo* roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. *Mol. Cell* **12,** 209–219 (2003).
23. Wolner, B., van Komen, S., Sung, P. & Peterson, C. L. Recruitment of the recombinational repair machinery to a DNA double-strand break in yeast. *Mol. Cell* **12,** 221–232 (2003).
24. Rodrigue, A. *et al.* Interplay between human DNA repair proteins at a unique double-strand break *in vivo*. *EMBO J.* **25,** 222–231 (2006).
25. West, S. C. Molecular views of recombination proteins and their control. *Nature Rev. Mol. Cell Biol.* **4,** 435–445 (2003).
26. Lee, M., Daniels, M. J. & Venkitaraman, A. R. Phosphorylation of BRCA2 by the Polo-like kinase Plk1 is regulated by DNA damage and mitotic progression. *Oncogene* **23,** 865–872 (2004).
27. Li, Z. *et al.* Alternative cyclin D1 splice forms differentially regulate the DNA damage response. *Cancer Res.* **70,** 8802–8811 (2010).
28. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463,** 899–905 (2010).
29. Bienvenu, F. *et al.* Transcriptional role of cyclin D1 in development revealed by a genetic-proteomic screen. *Nature* **463,** 374–378 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the members of the Sicinski lab for help and advice, N. Li for help with initial experiments, S Panyim and Y. Geng for discussions and reading the manuscript, P. Nakatani for pOZ-FH-N construct, M. Jasin for DR-GFP system, D. Bulavin and E. Appella for anti-phospho-CDC25A antibodies, A. Smogorzewska for anti-FANCI antibody, A. Venkitaraman and M. Lee for GST–BRCA2 fragments, and DFCI Confocal and Light Microscopy Core Facility for assisting with confocal microscopy. This work was supported by R01 CA083688, P01 CA080111 and P01 CA109901 grants from NIH (to P.S.). S.J. is supported by The Thailand Research Fund MRG5280248, W.M. by Foundation for Polish Science, Y.E.W. through the CCCB and the Dana-Farber Strategic Plan Initiative, A.B.C. and T.A.K. by Project Z01 ES065089 (to T.A.K.) from the Division of Intramural Research of the NIH, NIEHS.

Author Contributions S.J. and P.S. designed the study, analysed the data and wrote the manuscript. S.J. performed most of the experiments with help from collaborators. Y.H. and D.M.L. contributed to DNA repair analyses. W.M. contributed *in vitro* protein binding analyses. J.E.E. and T.G. performed mass spectrometry; J.E.E. analysed mass spec data with S.P.G. L.B., F.B., A.Z., T.v.H helped with experiments. Y.E.W. M.C. and J.Q. performed computational analyses of interactors. A.B.C. and T.A.K. contributed DNA mismatch repair analyses. B.X. helped with BRCA2 analyses.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.S. (peter_sicinski@dfci.harvard.edu).

doi:10.1038/nature09980

Structure and function of a membrane component SecDF that enhances protein export

Tomoya Tsukazaki[1]*, Hiroyuki Mori[2]*, Yuka Echizen[1], Ryuichiro Ishitani[1], Shuya Fukai[3,4], Takeshi Tanaka[5], Anna Perederina[6], Dmitry G. Vassylyev[6], Toshiyuki Kohno[5], Andrés D. Maturana[7], Koreaki Ito[8] & Osamu Nureki[1]

Protein translocation across the bacterial membrane, mediated by the secretory translocon SecYEG and the SecA ATPase[1–4], is enhanced by proton motive force[5,6] and membrane-integrated SecDF[7–9], which associates with SecYEG. The role of SecDF has remained unclear, although it is proposed to function in later stages of translocation as well as in membrane protein biogenesis[4,10–13]. Here, we determined the crystal structure of *Thermus thermophilus* SecDF at 3.3 Å resolution, revealing a pseudo-symmetrical, 12-helix transmembrane domain belonging to the RND superfamily and two major periplasmic domains, P1 and P4. Higher-resolution analysis of the periplasmic domains suggested that P1, which binds an unfolded protein, undergoes functionally important conformational changes. *In vitro* analyses identified an ATP-independent step of protein translocation that requires both SecDF and proton motive force. Electrophysiological analyses revealed that SecDF conducts protons in a manner dependent on pH and the presence of an unfolded protein, with conserved Asp and Arg residues at the transmembrane interface between SecD and SecF playing essential roles in the movements of protons and preproteins. Therefore, we propose that SecDF functions as a membrane-integrated chaperone, powered by proton motive force, to achieve ATP-independent protein translocation.

The functional importance of SecDF in protein translocation was previously shown *in vivo*[7,9,14]: the SecD- and SecF-deficient *Escherichia coli* mutant *secD1* (producing negligible levels of SecD and SecF and exhibiting cold-sensitivity (Cs) for growth) is severely defective in protein export at any temperature. In *E. coli*, SecD and SecF (EcSecD and EcSecF respectively) are encoded by the *secD* operon (*yajC-secD-secF*) and form a tight complex[8]. Each protein contains six transmembrane (TM) segments, belonging to the resistance nodulation and cell division (RND) superfamily, and also a large periplasmic domain[15,16].

SecDF from *Thermus thermophilus* HB8[17] is composed of a single polypeptide chain (TtSecDF, molecular mass 80.5 kDa) containing 12 predicted TM regions and six periplasmic regions (P1–P6), among which P1 ($TtSecDF_{31–263}$) and P4 ($TtSecDF_{469–559}$) are large enough to form distinct domains (Supplementary Fig. 1). We solved the crystal structure of full-length TtSecDF at 3.3 Å resolution (Fig. 1a–c, f, Supplementary Fig. 2 and Supplementary Table 1), as well as the crystal structure of P1 ($TtSecDF_{36–263}$) at 2.6 Å resolution and the nuclear magnetic resonance (NMR) solution structure of P4 ($TtSecDF_{481–557}$) (Fig. 1d, e and Supplementary Table 2). The TtSecDF structure contains twelve TM α-helices with both termini facing the cytoplasm and P1 and P4 protruding into the periplasm (Fig. 1a, b). The N-terminal (TM1–6) and C-terminal (TM7–12) halves are assembled pseudo-symmetrically, with TM4 and TM10 forming the primary interface between SecD and SecF (Fig. 1c). The TM helices are membrane-embedded, except for the ~10 Å extensions of TM2 and TM8 that tether P1 and P4 respectively (Fig. 1a). The tilted TMs 4, 5, 6, 10, 11 and 12 are curved near the cytoplasmic surface.

The NMR structure of P4 revealed a ferredoxin-like fold, consisting of four β-strands flanked by two α-helices (Fig. 1e). P1, protruding by ~40 Å from the membrane, is divided into head (~25 Å periplasmic protrusion) and base subdomains (Fig. 1a, d). The base subdomain of P1 (residues 36–111 and 251–263) is structurally homologous to P4; both form pseudo-symmetrical, eight-stranded, anti-parallel β-sheets that cover the TM region (Fig. 1a, b). The crystal structure of isolated P1 (called the I form) revealed a ~120° rigid-body rotation of its head subdomain towards the periplasmic space, as compared to the orientation observed for P1 in full-length SecDF (called the F form) (Supplementary Fig. 3). Thus, SecDF can assume at least two different conformations, F and I, with distinct P1 configurations (Fig. 1f, g). For this conformational change, the two loops connecting the head and the base could act as a hinge (Fig. 1a, d), allowing the head to swing around the base. We built a model of the 'full-length I form' by superimposing the base subdomain of the isolated P1 structure onto that of the full-length SecDF. The *in vivo* occurrence of the two conformations was supported by disulphide crosslinking experiments, which further suggested that the conformational transition of SecDF is functionally important (Supplementary Fig. 4a–e and Supplementary Discussion).

Despite the vital importance of SecDF *in vivo*, its specific role in translocation has not been defined. We examined whether SecDF functions in the steps subsequent to the SecA- and ATP-dependent initiation of translocation. A pro-outer membrane protein A (pro-OmpA) translocation intermediate was generated by subjecting ^{35}S-labelled pro-OmpA(L59)[18] to translocation reactions in the presence of SecA and ATP, with either wild-type inverted membrane vesicles (IMVs) or SecD- and SecF-deficient IMVs from the *secD1* (Cs) mutant. This intermediate possessed a 59-residue disulphide-linked loop acting as an obstacle to translocation through the Sec translocon (Fig. 2b). ATP was then depleted from the intermediate-bearing IMVs by glucose-hexokinase reaction cycles (Fig. 2a, lanes 1 and 6). The addition of dithiothreitol to the wild-type IMVs restored translocation that could be supported by a proton motive force (PMF) even in the absence of ATP (compare Fig. 2a, lanes 2 and 4). In contrast, the *SecD1* IMVs did not support any further translocation of the intermediate, regardless of the presence of a PMF (Fig. 2a, lane 7). Instead, these IMVs tended to lose the intermediate (Fig. 2a, lanes 7 and 8), probably owing to backsliding of the polypeptide[19]. In the presence of excess ATP, the completion of pro-OmpA(L59) translocation was not affected by either the *secD1* mutation or the lack of PMF (Fig. 2a, lanes 13, 15, 18 and 20). These results show that SecDF is required for the later, ATP-independent steps of translocation driven by the PMF, whereas the PMF is thought to accelerate even the initial stage of translocation (Supplementary Discussion). In addition, IMVs bearing inactive

[1]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan. [2]Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto 606-8507, Japan. [3]Structural Biology Laboratory, Life Science Division, Synchrotron Radiation Research Organization, The University of Tokyo, Tokyo 113-0032, Japan. [4]Laboratory of Macromolecular Complexes, Center for Structural Biology of Challenging Proteins, The University of Tokyo, Tokyo 113-0032, Japan. [5]Mitsubishi Kagaku Institute of Life Sciences, Machida-shi, Tokyo 194-8511, Japan. [6]Department of Biochemistry and Molecular Genetics, Schools of Medicine and Dentistry, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. [7]Bioengineering Department, Nagaoka University of Technology, Niigata 940-2188, Japan. [8]Kyoto Sangyo University, Kita-ku, Kyoto 603-8555, Japan.
*These authors contributed equally to this work.

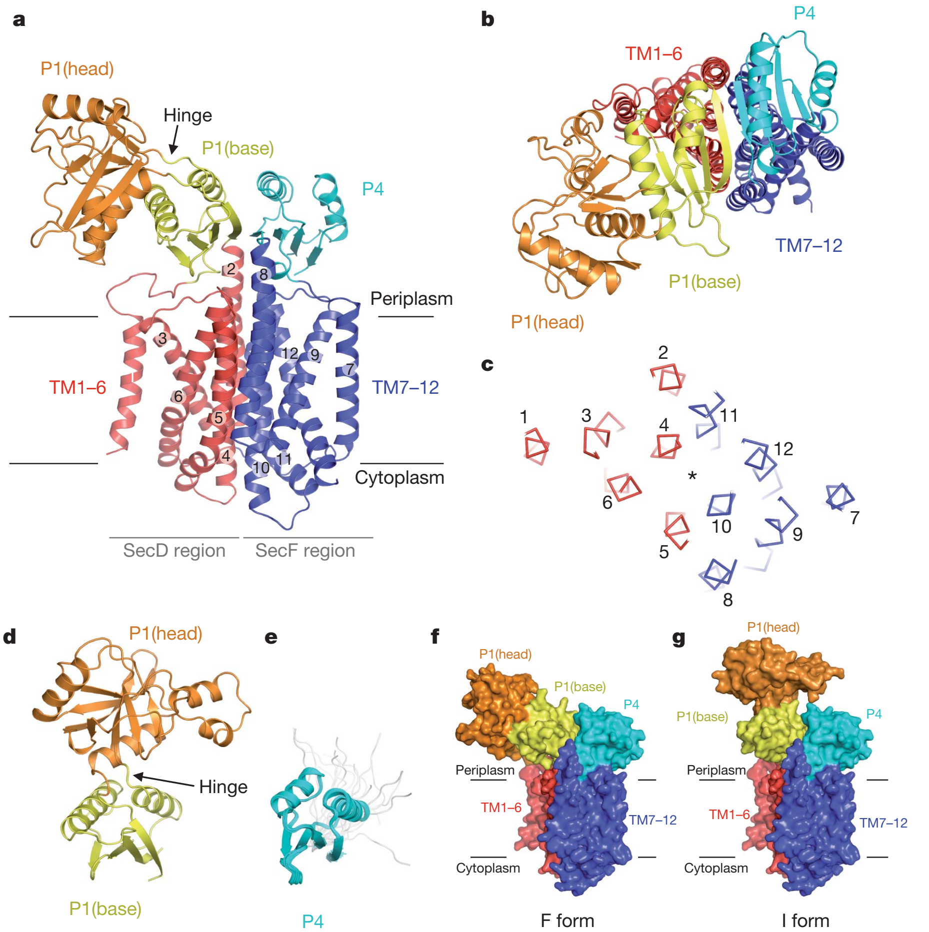


Figure 1 | Structures of *T. thermophilus* SecDF. **a**, **b**, The crystal structure of full-length SecDF viewed from the membrane side (**a**) and the periplasmic side (**b**). **c**, TtSecDF cross-sectioned at the middle of the TM, viewed from the periplasm. The asterisk indicates the pseudo-symmetrical axis. TMs are numbered. **d**, Crystal structure of the P1 domain. **e**, NMR structure of the P4 domain after twenty superimpositions. The disordered regions are shown in grey. **f**, Crystal structure of full-length SecDF, F form. **g**, Crystal structure of full-length SecDF, I form. The base subdomain of isolated P1 was docked onto that of the F form, as shown in (**f**).

SecDF variants (described below) did not support the PMF-dependent completion of pro-OmpA translocation (Fig. 2c).

To gain structural insights into the PMF-dependence of SecDF function, we compared its structure with that of AcrB[20], an RND superfamily proton/multi-drug antiporter[16] (Supplementary Fig. 5). AcrB forms a homotrimer, whereas SecDF is monomeric. Although the TM segments of SecDF and AcrB share low sequence identity of 15% (Supplementary Fig. 6), the structures of their TM regions are similar, yielding a root mean square deviation of ~2.7 Å for the Cα atoms of the TM helices. By contrast, the structures and functions of the periplasmic regions are different between SecDF and AcrB. The TM region of AcrB is thought to participate in proton transport and contains several conserved, charged residues important for drug export activity, such as Asp 407, Asp 408, Lys 940 and Arg 971[21] (Supplementary Fig. 5e). Asp 407, Thr 978 and Arg 971 in AcrB have structural counterparts in TtSecDF: Asp 340, Thr 675 and Arg 671 respectively (Fig. 3a and Supplementary Fig. 5e). The conserved SecDF residues are clustered at the TM interfaces between SecD and SecF and in the periplasmic base region underneath the head (Supplementary Fig. 7). We also note that Asp 637 is a highly conserved, membrane-embedded charged residue. Complementation tests indicated that the Asp519Asn mutation in EcSecD as well as the Asp213Asn and Arg247Met mutations in EcSecF (Fig. 3a) completely abolished SecDF activity and conferred some dominant-negative phenotypes (Fig. 3b). Thus, these conserved charged residues are crucial for SecDF function, consistent with the hypothesis that SecDF conducts protons through the conserved TM region.

The halophilic marine bacteria *Vibrio* spp. use a sodium ion (Na^+) gradient instead of PMF for some cellular processes[22]. *Vibrio alginolyticus* has two sets of *secDF* genes encoding SecDF-1 and SecDF-2 complexes respectively. When VaSecDF-1 from *V. alginolyticus* 138-2 was expressed in the *E. coli secD1* (Cs) mutant, the addition of NaCl,

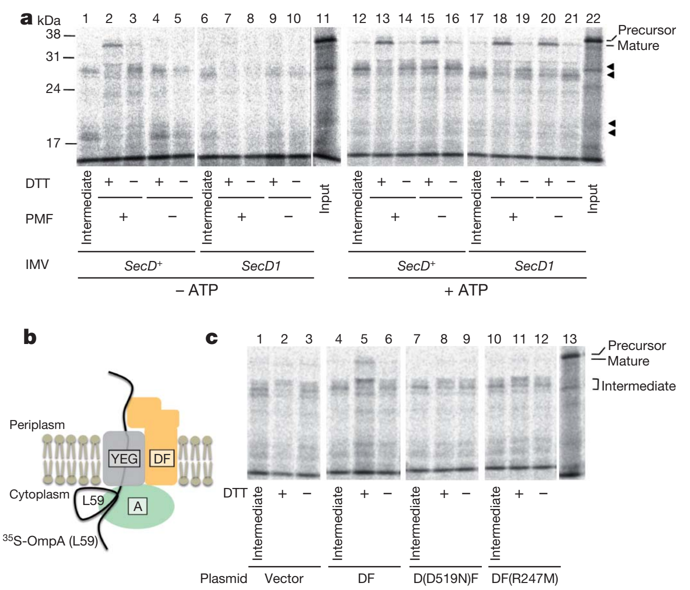


Figure 2 | SecDF-dependent translocation completion. **a**, Identification of a SecDF- and PMF-dependent translocation step. SecDF-deficient (*secD1*) IMVs were incubated with ^{35}S-labelled pro-OmpA(L59) to generate translocation intermediates. Protein translocation was continued in the presence or absence of ATP, PMF and dithiothreitol (DTT) before SDS-PAGE and phosphor imaging. Arrowheads indicate translocation intermediates. **b**, Schematic depiction of the translocation intermediate of pro-OmpA(L59). A, SecA; DF, SecDF; YEG, SecYEG. **c**, Completion of pro-OmpA(L59) translocation using IMVs from the *secD1* (Cs) mutant expressing no additional protein (vector), *E. coli* SecDF or the SecDF derivatives SecD(D519N)F and SecDF(R247M).

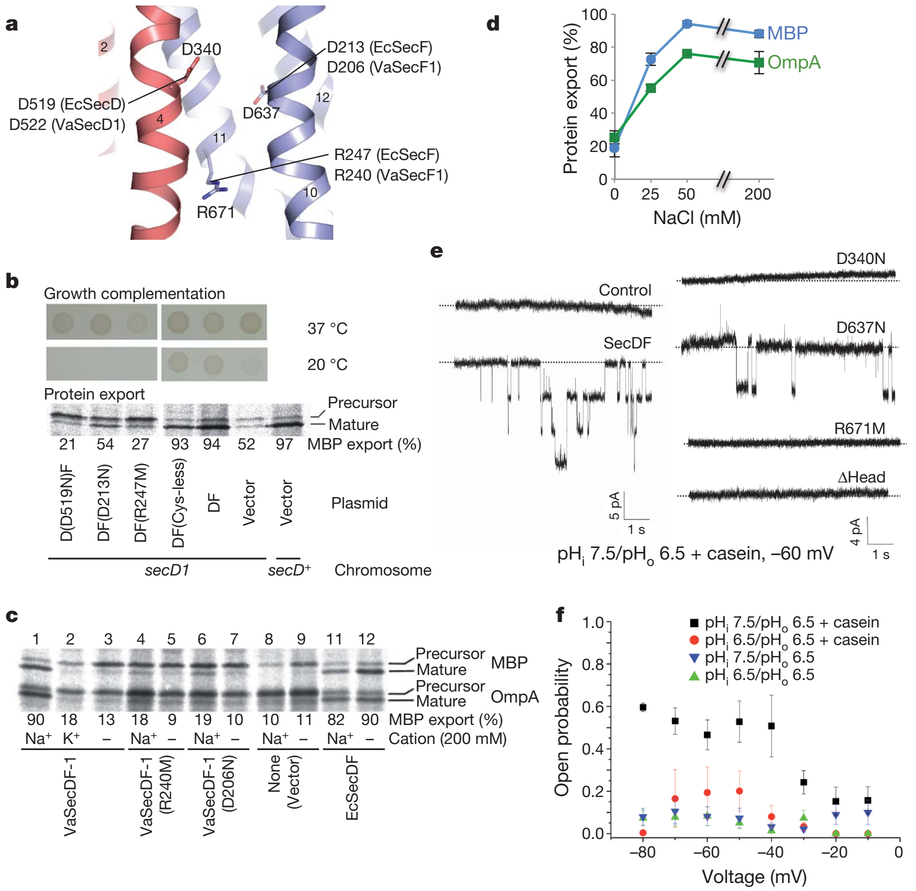


Figure 3 | Functional charged residues and proton conduction of SecDF. **a**, Functionally important, conserved residues in the TM regions. The TMs are numbered. **b**, Complementation of the *secD1* (Cs) growth and protein export defects by SecD and SecF mutants. Assay, SDS-PAGE and phosphor imaging as in Fig. 2. MBP, maltose-binding protein. **c**, Na^+-dependent protein export by VaSecDF-1. **d**, Na^+-dependence of protein export by VaSecDF-1 ($n = 3$). **e**, Single channel currents recorded by patch clamp in membrane patches excised from *E. coli* spheroplasts containing TtSecDF (left panel) or its mutants D340N, D637N, R671M and ΔHead (right panel) and effects of a pH gradient and casein. **f**, The channel open probability in experiments shown in (**e**) ($n = 6$ with casein, $n = 4$ without casein). Error bars indicate s.e.m.

but not KCl, to the medium restored protein export (Fig. 3c, lanes 1–3). Export was dependent on the extracellular Na^+ concentration up to 50 mM (Fig. 3d) but such Na^+-dependence was not observed with EcSecDF (Fig. 3c, lanes 11 and 12). Thus, VaSecDF-1 facilitates protein export using a Na^+ gradient across the membrane. The Arg240Met and Asp206Asn mutations in VaSecF-1, corresponding to EcSecF Arg247Met and Asp213Asn respectively (Fig. 3a), compromised the Na^+-dependent protein export activity (Fig. 3c, lanes 4–7). These results provide physiological evidence for cation-coupled protein translocation by SecDF.

To verify that SecDF utilizes the PMF and thus conducts protons, we performed inside-out patch clamp experiments[23] using *E. coli* giant spheroplasts[24] containing TtSecDF. Current recordings revealed that spheroplasts containing TtSecDF underwent transient channel openings under both symmetrical and asymmetrical pH conditions, whereas spheroplasts without TtSecDF did not (Supplementary Fig. 8a). The channel activities were markedly enhanced by the imposition of a pH gradient as well as by the addition of a P1-interacting unfolded protein, casein, to the pipette (periplasmic side) solution (Fig. 3e left panel, Supplementary Fig. 8a–c, Supplementary Discussion). In addition, casein and acidic extracellular conditions increased the probability of opening of the SecDF-dependent ion channel (Fig. 3f). Thus, ion-conduction can be regulated by a proton gradient and by the binding of an unfolded protein to SecDF. The use of a proton-specific fluorescent probe, BCECF-AM (2′,7′-bis-(2-carboxyethyl)-5-(and-6)-carboxyfluorescein, acetoxymethyl ester), confirmed SecDF-dependent proton import (Supplementary Fig. 9 and Supplementary Discussion). The TtSecDF mutations Asp340Asn and Arg671Met, but not Asp637Asn, abolished ion channel activity (Fig. 3a and 3e right panel), indicating that Asp 340 and Arg 671 in the TM region are essential for proton conduction and might be protonated transiently. Asp 637, which is sequestered from the putative main proton pathway in the TM region of both AcrB and SecDF (Supplementary Fig. 5e), might function in subsequent conformational changes that are required for the enhancement of protein translocation.

In addition to the TM residues, the P1 head domain is a critical element for proton transport. TtSecDF(Δ112–248), which lacks the head, did not show any signals in the patch clamp assay (Fig. 3e right panel) and the corresponding EcSecDF mutant was defective in protein export (Supplementary Fig. 4f). The conformational flexibility of P1 is also required for proton conduction because the TtSecDF double cysteine mutant, Leu106Cys/Leu243Cys (see Supplementary Fig. 4a, b), lacked channel activity in four of the six membrane patches examined. The variable results could be explained by incomplete crosslinking (more than 20%, data not shown). We think it is likely that the putative hinge motion of the P1 domain is coupled with both proton transport and protein export. Taken together, we propose that proton flow is the driving force for P1 domain movement and for the consequent enhancement of protein translocation by SecDF.

We have shown here that SecDF and PMF are required for the post-initiation mode of translocation, which can occur in the absence of ATP and SecA. This function depends on the ability of the periplasmic P1 domain to interact with a substrate and to undergo a structural transition between the I and F forms, which are likely to be in equilibrium. The F state of SecDF may place the tilted P1 head above the translocon pore, enabling it to capture an emerging preprotein (Fig. 4a). The preprotein-bearing F form could then return to the I configuration (Fig. 4b), preventing the backward movement of the substrate and driving its forward movement. The release of the bound preprotein from SecDF and the subsequent I to F conversion may be coupled to proton flow (Fig. 4c). These action cycles will eventually lead to the completion of translocation, in which the substrate is released from the translocon. As the bacterial periplasm lacks ATP, SecDF may utilize the PMF to drive its conformational transition and delivery of substrates. The Asp and Arg residues in the TM region of SecDF could serve as putative proton acceptors in the proton relay pathway (Fig. 4c). The corresponding residues of AcrB (Supplementary Fig. 5e) have been proposed to participate in proton conduction with their side chains assuming different configurations among the asymmetric protomers,

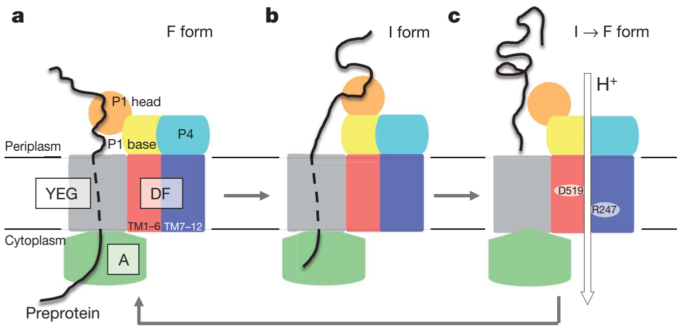


Figure 4 | A working model of the PMF-driven translocation enhancement by SecDF. **a**, F form, capturing state. **b**, I form, holding state. **c**, I to F transition and substrate-releasing state. The two essential charged residues of EcSecDF are highlighted. SecDF is coloured as in Fig. 2. SecYEG, grey; SecA, green; preprotein, black line; proton movement, white arrow. See the main text discussion for details.

which presumably exist in different protonation states[20]. Likewise, protonation of the key charged residues of SecDF could induce the twisting of TM4 and TM10. This would be transmitted to the conserved P1-TM4 linker region (Supplementary Fig. 7) and would trigger the conformational transition of the P1 head subdomain.

Although direct evidence for the PMF-dependent conformational transition of SecDF will await further structural and functional studies, we have shown that SecDF is a component of the Sec machinery that utilizes the PMF to complete protein translocation after the ATP-dependent function of SecA.

METHODS SUMMARY

X-ray diffraction analysis of TtSecDF was described previously[17]. The initial phases were determined by the single-wavelength anomalous dispersion method. The initial model was manually built, using the structures of the TM regions of AcrB and the separately-determined P1 and P4 domain structures as references (Supplementary Methods). The model was substantially improved by using the zonal scaling[25] and methionine-marking[26] methods and was finally refined to $R_{work} = 29.8\%$ and $R_{free} = 31.9\%$ at 3.3 Å resolution.

To monitor the later steps of translocation, ^{35}S-labelled pro-OmpA(L59) and IMVs, prepared as described previously[18,27], were incubated to form translocation intermediates which were then incubated under several conditions and treated with proteinase K. The translocation status of ^{35}S-pro-OmpA was examined by phosphor imaging after SDS-polyacrylamide gel electrophoresis (SDS–PAGE).

For the complementation test, *E. coli secD1* (Cs) mutant cells carrying plasmids encoding HA-tagged EcSecD and EcSecF mutants were spotted onto LB agar plates (pH adjusted to ~7.5) and incubated at 20 °C. The efficiency of protein export *in vivo* was assessed by the ^{35}S-methionine pulse-labelling procedures.

To monitor the protein transport activity of *Vibrio* SecDF-1, *E. coli secD1* cells expressing VaSecD-1 and VaSecF-1 were pulse-labelled in the presence or absence of Na^+. To detect single channel activity, electric currents were measured with inside-out membrane patches excised from TtSecDF-containing *E. coli* giant spheroplasts, using an Axopatch 200B amplifier (Axon CNS, Molecular Devices).

Received 11 January 2010; accepted 9 March 2011.
Published online 11 May 2011.

1. van den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* **427,** 36–44 (2004).
2. Zimmer, J., Nam, Y. & Rapoport, T. A. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455,** 936–943 (2008).
3. Tsukazaki, T. *et al.* Conformational transition of Sec machinery inferred from bacterial SecYE structures. *Nature* **455,** 988–991 (2008).
4. du Plessis, D. J., Nouwen, N. & Driessen, A. J. The Sec translocase. *Biochim. Biophys. Acta* **1808,** 851–865 (2011).
5. Driessen, A. J. & Wickner, W. Proton transfer is rate-limiting for translocation of precursor proteins by the *Escherichia coli* translocase. *Proc. Natl Acad. Sci. USA* **88,** 2471–2475 (1991).
6. Shiozuka, K., Tani, K., Mizushima, S. & Tokuda, H. The proton motive force lowers the level of ATP required for the *in vitro* translocation of a secretory protein in *Escherichia coli. J. Biol. Chem.* **265,** 18843–18847 (1990).
7. Pogliano, J. A. & Beckwith, J. SecD and SecF facilitate protein export in *Escherichia coli. EMBO J.* **13,** 554–561 (1994).
8. Sagara, K., Matsuyama, S. & Mizushima, S. SecF stabilizes SecD and SecY, components of the protein translocation machinery of the *Escherichia coli* cytoplasmic membrane. *J. Bacteriol.* **176,** 4111–4116 (1994).
9. Hand, N. J., Klein, R., Laskewitz, A. & Pohlschroder, M. Archaeal and bacterial SecD and SecF homologs exhibit striking structural and functional conservation. *J. Bacteriol.* **188,** 1251–1259 (2006).
10. Matsuyama, S., Fujita, Y. & Mizushima, S. SecD is involved in the release of translocated secretory proteins from the cytoplasmic membrane of *Escherichia coli. EMBO J.* **12,** 265–270 (1993).
11. Economou, A., Pogliano, J. A., Beckwith, J., Oliver, D. B. & Wickner, W. SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF. *Cell* **83,** 1171–1181 (1995).
12. Arkowitz, R. A. & Wickner, W. SecD and SecF are required for the proton electrochemical gradient stimulation of preprotein translocation. *EMBO J.* **13,** 954–963 (1994).
13. Duong, F. & Wickner, W. The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. *EMBO J.* **16,** 4871–4879 (1997).
14. Nouwen, N., Piwowarek, M., Berrelkamp, G. & Driessen, A. J. The large first periplasmic loop of SecD and SecF plays an important role in SecDF functioning. *J. Bacteriol.* **187,** 5857–5860 (2005).
15. Pogliano, K. J. & Beckwith, J. Genetic and molecular characterization of the *Escherichia coli* secD operon and its products. *J. Bacteriol.* **176,** 804–814 (1994).
16. Tseng, T. T. *et al.* The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *J. Mol. Microbiol. Biotechnol.* **1,** 107–125 (1999).
17. Tsukazaki, T. *et al.* Purification, crystallization and preliminary X-ray diffraction of SecDF, a translocon-associated membrane protein, from Thermus thermophilus. *Acta Crystallogr. F* **62,** 376–380 (2006).
18. Uchida, K., Mori, H. & Mizushima, S. Stepwise movement of preproteins in the process of translocation across the cytoplasmic membrane of *Escherichia coli. J. Biol. Chem.* **270,** 30862–30868 (1995).
19. Schiebel, E., Driessen, A. J., Hartl, F. U. & Wickner, W. $\Delta\mu_{H}+$ and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell* **64,** 927–939 (1991).
20. Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T. & Yamaguchi, A. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature* **443,** 173–179 (2006).
21. Seeger, M. A., von Ballmoos, C., Verrey, F. & Pos, K. M. Crucial role of Asp408 in the proton translocation pathway of multidrug transporter AcrB: evidence from site-directed mutagenesis and carbodiimide labeling. *Biochemistry* **48,** 5801–5812 (2009).
22. Häse, C. C. & Barquera, B. Role of sodium bioenergetics in *Vibrio cholerae. Biochim. Biophys. Acta* **1505,** 169–178 (2001).
23. Sasaki, M., Takagi, M. & Okamura, Y. A voltage sensor-domain protein is a voltage-gated proton channel. *Science* **312,** 589–592 (2006).
24. Hattori, M. *et al.* Mg^{2+}-dependent gating of bacterial MgtE channel underlies Mg^{2+} homeostasis. *EMBO J.* **28,** 3602–3612 (2009).
25. Vassylyev, D. G. *et al.* Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448,** 163–168 (2007).
26. Inaba, K. *et al.* Crystal structure of the DsbB-DsbA complex reveals a mechanism of disulfide bond generation. *Cell* **127,** 789–801 (2006).
27. Matsuo, E., Mori, H., Shimoike, T. & Ito, K. Syd, a SecY-interacting protein, excludes SecA from the SecYE complex with an altered SecY24 subunit. *J. Biol. Chem.* **273,** 18835–18840 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Akiyama, R. Suno, Y. Morimoto, T. Minamino, K. Namba, K. Inaba, M. Hattori and H. Nishimasu for suggestions; T. Sakamoto and A. Kurabayashi for assistance with sample preparation; R. Yamasaki, M. Sano, K. Mochizuki, K. Yoshikaie, K. Imayoshi and T. Adachi for technical support; M. Homma and S. Kojima for providing the *Vibrio* genomic DNA; the beamline staff members at BL41XU of SPring-8 (Hyogo, Japan) and at NW12 of KEK PF-AR (Tsukuba, Japan) for technical help during data collection and M. Ibba for comments on our manuscript. This work was supported by a Grant-in-Aid for Scientific Research (S) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to O.N., by a CREST grant from JST to K.I., by a BIRD grant from JST to H.M. and R.I., by a grant for the National Project on Protein Structural and Functional Analyses to O.N., by NIH grants to D.G.V. and by grants from MEXT to T.Tsukazaki, H.M., R.I., S.F. and K.I.

Author Contributions T.Tsukazaki performed the structural determination and the biochemical experiments with SecDF. H.M. performed the functional analyses of SecDF. Y.E. solved the crystal structure of the SecDF P1 domain and assisted with the functional analysis of SecDF. R.I., S.F., D.G.V. and O.N. assisted with the structural determination. A.D.M. performed patch clamp and pH fluorescence experiments. T.Tanaka and T.K. solved the structure of the P4 domain by NMR. A.P. and D.G.V. assisted with the crystallization and data collection of SecDF. All authors discussed the results and commented on the manuscript. O.N. and K.I. supervised the work and wrote and edited the manuscript.

Author Information The coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 3AQP for the entire TtSecDF protein and 3AQO for the P1 domain. The PDB and BMRB codes for the deposited P4 domain are 2RRN and 11426 respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to O.N. (nureki@ims.u-tokyo.ac.jp) or K.I. (kito@cc.kyoto-su.ac.jp).

CORRECTIONS & AMENDMENTS

ERRATUM

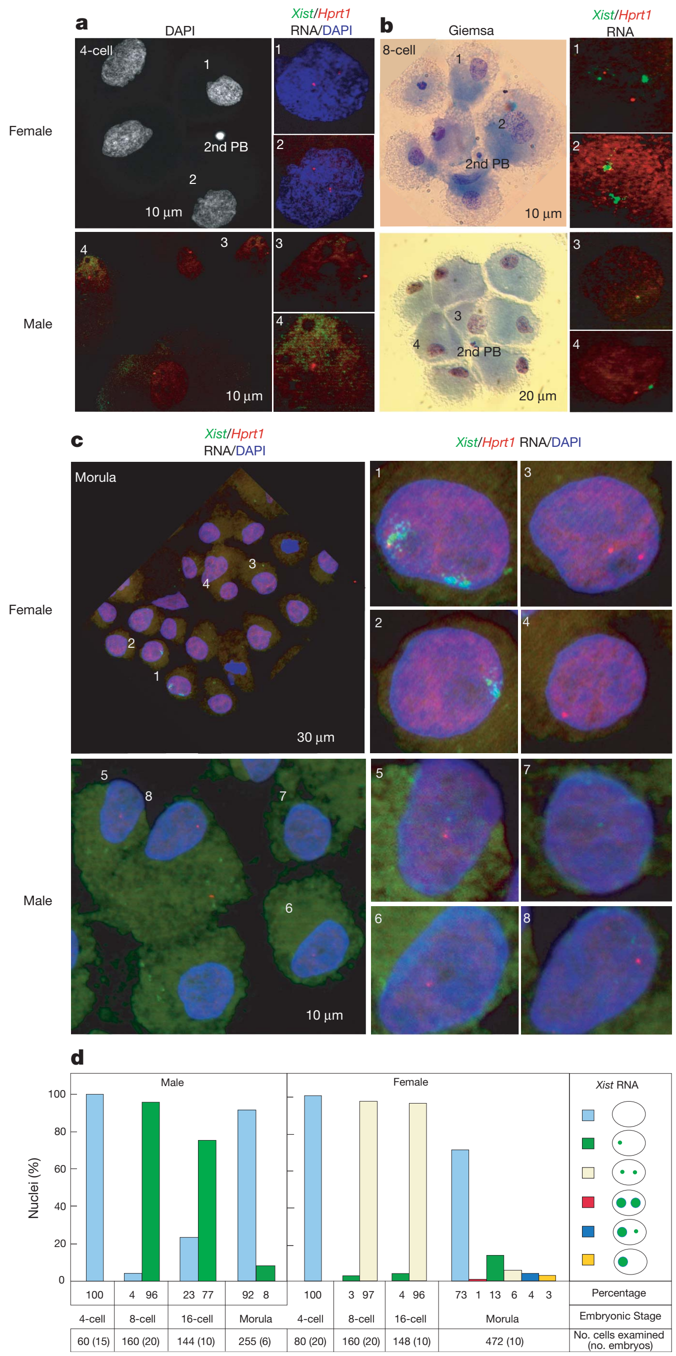
doi:10.1038/nature10184

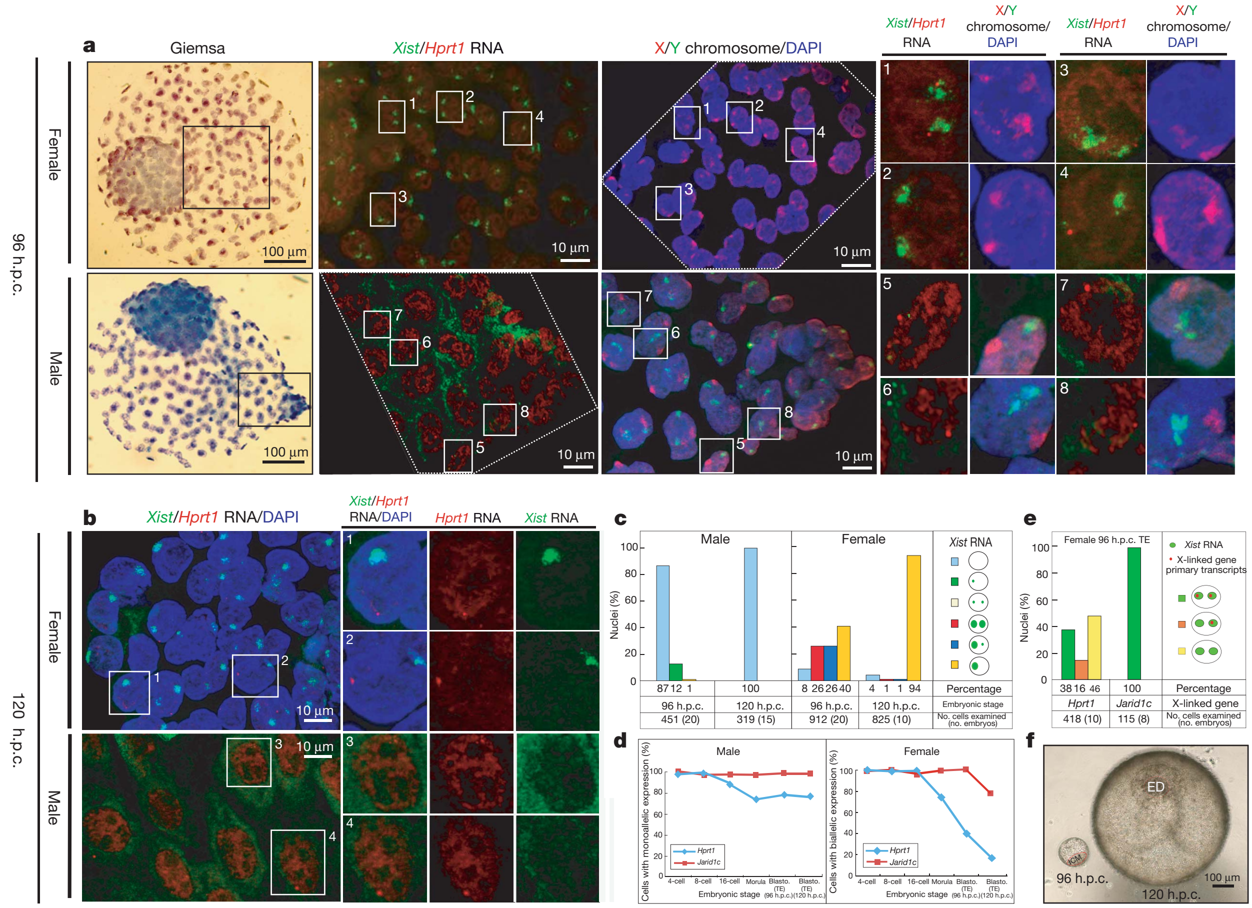
Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development

Ikuhiro Okamoto, Catherine Patrat, Dominique Thépot, Nathalie Peynot, Patricia Fauque, Nathalie Daniel, Patricia Diabangouaya, Jean-Philippe Wolf, Jean-Paul Renard, Véronique Duranthon & Edith Heard

Nature **472**, 370–374 (2011)

In this Letter, Figs 1 and 2 were inadvertently printed at low resolution. The corrected Figs 1 and 2 are presented here and overleaf, and these have been replaced in the HTML and PDF version of the manuscript.




a
Giemsa
Xist/Hprt1 RNA
X/Y chromosome/DAPI
96 h.p.c.
Female
Male
100 µm
10 µm
b
Xist/Hprt1 RNA/DAPI
Hprt1 RNA
Xist RNA
120 h.p.c.
c
Male
Female
Xist RNA
Nuclei (%)
Percentage
Embryonic stage
96 h.p.c.
120 h.p.c.
451 (20)
319 (15)
912 (20)
825 (10)
No. cells examined (no. embryos)
d
Cells with monoallelic expression (%)
Cells with biallelic expression (%)
Hprt1
Jarid1c
Embryonic stage
e
Female 96 h.p.c. TE
X-linked gene primary transcripts
X-linked gene
418 (10)
115 (8)
f
ED
ICM
96 h.p.c.
120 h.p.c.

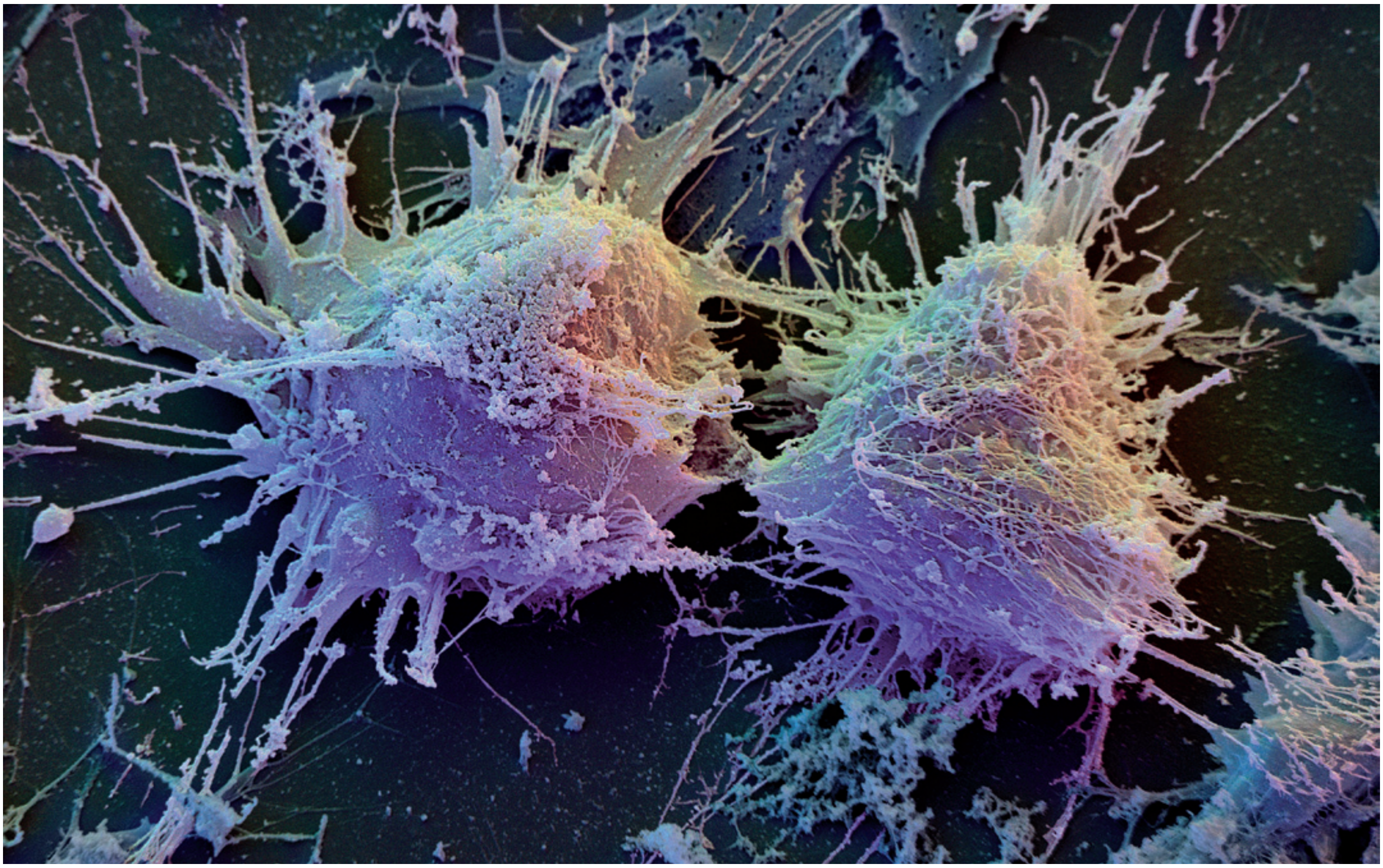
CAREERS

CAREER TOOL KIT Access key resources to craft a successful career **go.nature.com/5lzhj6**

CAREERS ADVICE FORUM Get expert advice on science careers issues **go.nature.com/lm1x4t**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

D. SCHARF/SCIENCE FACTION/CORBIS



Mastering the growth of human embryonic stem cells, including by methods that use mouse fibroblasts (pictured), continues to spur research opportunities.

BIOMEDICAL RESEARCH

Growing with the flow

A US court case is making stem-cell researchers nervous. But the field continues to thrive.

BY MEREDITH WADMAN

As Renée Reijo Pera prepared for a radio interview one late-April morning in her office in Palo Alto, California, she was savouring two pieces of good news. The day before, the faculty senate had unanimously approved a new PhD programme in stem-cell biology and regenerative medicine at Stanford University's medical school. Reijo Pera will be directing the programme when it starts up this autumn. And just that morning, the US Court of Appeals for the District of Columbia Circuit — one level below the US Supreme Court — had thrown out a lower-court injunction. The injunction had blocked the US National Institutes of Health (NIH) from funding research that used human embryonic stem (ES) cells for 17 days last summer — and threatened to derail the funding for good. "I remember saying that good news often seems to come in waves," says Reijo Pera. But, as well as being relieved about the court ruling, she says, she was also wary. "If you're in this field you become a little leery of getting your hopes up. I wanted to know more, to make sure that I had heard correctly."

Reijo Pera had heard correctly, but she was right to be on guard. The ruling is unlikely to be the last chapter in this closely watched case (see *Nature* **473,** 15; 2011). What is more, an ultimate victory in the case, which pitted two scientist activists against the Department of Health and Human Services, would not prevent future presidents or Congresses from acting anew to limit government funding for the research. Surveys suggest that a majority of Americans approve of human-ES-cell research, but there is a vocal opposition that objects to days-old human embryos being destroyed to obtain cell lines, and is willing to wage legal battles to stop research on them. So, although the NIH continues to fund the research — including an estimated US$125 million this year — it does not do so with the legal and political security that exists, for ▶

instance, in the United Kingdom.

Young scientists considering or embarking on careers in stem-cell biology have to come to terms with the uncertainty that surrounds human-ES-cell research funding. But although funding issues can make the career road bumpy, the exploding number of opportunities in the field means that it can still be negotiated with forethought and careful planning.

BE WILLING TO MOVE

Students choosing a graduate institution or postdoc position, or people seeking a faculty position or industry job (see 'Industry's stem-cell possibilities'), should consider looking in the jurisdictions most hospitable to stem-cell research. In the United States, four states — California, Connecticut, Maryland and New York — collectively funded more human-ES-cell research than the NIH did between 2005 and 2009, and their funding equalled or exceeded the NIH's in 2008 and 2009 (R. N. Karmali, N. M. Jones and A. D. Levine *Nature Biotechnol.* **28,** 1246–1248; 2010). Positions in these states therefore often have access to funding from sources other than NIH grants.

"That's particularly important if you're a young stem-cell scientist in this field," says Aaron Levine, an assistant professor of public policy at the Georgia Institute of Technology in Atlanta. Researchers can thrive in "places where you have alternative streams of funding to protect you if NIH funding gets interrupted again", he says.

Since he arrived at the University of California, San Francisco, in 2008 to begin his second postdoc, Dutch scientist Dennis Van Hoof has been insulated from the vagaries of US stem-cell politics. His buffer is a combination of private foundation grants and infrastructure support from the California Institute for Regenerative Medicine (CIRM) in San Francisco, which is dispersing \$3 billion on stem-cell research funded by California taxpayers.

CIRM has handed out more than \$428 million in human-ES-cell science awards. In addition, it has funded more than 200 research-training and career-development grants, collectively worth about \$386 million, since 2006. It has also contributed to new buildings. Van Hoof, who himself has type-I diabetes, is working to produce insulin-secreting pancreatic beta cells from human ES cells to replace those that no longer function in people with diabetes. He works in a brand-new stem-cell research building funded by CIRM, and his grant support has come largely from the Juvenile Diabetes Research Foundation and the Leona M. & Harry B. Helmsley Charitable Trust, both based in New York. He has never applied for an NIH grant and he intends to keep it that way. "If you have an idea for a grant proposal, federal funding would definitely be second choice," he says.

"If stem cells are going to move out of the lab, there will be lots of need for engineers to produce a large number of identical cells."
Aaron Levine

However, apart from a fairly exclusive stable of labs well supported by private funding, private donors don't begin to approach what the NIH can offer. Scientists often use private money to seed projects for which they hope eventually to win NIH support. And not all private foundations embrace human-ES-cell research. The American Heart Association in Dallas, Texas, for instance, has steered clear of funding it because of the ethical and political controversy involved.

That kind of caution should disappear with time, especially if human ES cells begin to produce results in the clinic, according to some researchers. "I think it will not be long before we see real clinical benefit from this research," argues Martin Pera, who on 1 June became programme leader of Stem Cells Australia at the University of Melbourne. "And once that happens, the opposing voices will diminish in influence."

GEORGIA TECH

SAFE HAVENS?

Stem-cell 'havens' are not always easy to find. Pera — who until last month directed the stem-cell research centre at the University of Southern California in Los Angeles — is returning to his native Australia at a time when Australian government support for the research has been scaled back by just over 70%. At the end of this month, the government will close the Australian Stem Cell Centre, which funded stem-cell research to the tune of Aus\$11 million (US\$12 million) in the last financial year. In its place, Stem Cells Australia will be funded at Aus\$21 million spread over seven years.

By contrast, at the University of Cambridge, UK, the Cambridge Stem Cell Initiative — an

BRANCHING OUT

Industry's stem-cell possibilities

Like academia, industry promises many opportunities for stem-cell researchers — but they bring their own uncertainties.

The opportunities come from the many companies working to develop adult stem-cell therapies. And in the field of human embryonic stem (ES) cells, businesses in the United States are somewhat insulated, by virtue of their private funding, from the political vagaries that dictate the livelihoods of researchers funded by the US National Institutes of Health (NIH).

But there are uncertainties. Businesses are not immune to tough economic circumstances. In February, for example, Pfizer announced the closure of its three-year-old Regenerative Medicine unit, which had been developing stem-cell therapies for cardiac disorders and cancer. Its demise eliminated 18 jobs in Cambridge, Massachusetts. The company's remaining regenerative-medicine work has been folded into a subsidiary in Cambridge, UK, now called Neusentis, where roughly 30 people still work on human ES cells. They focus on topics ranging from creating cell lines for drug discovery to developing cell therapies for diseases such as macular degeneration, a leading cause of blindness.

Before being put off by Pfizer's downsizing, job-seekers should remember that, when it launched Pfizer Regenerative Medicine, the company put itself out on the most vulnerable edge of the human-ES-cell research spectrum: bringing cures to the clinic. "Cell therapy is a very risky enterprise for big investment," says John McNeish, former executive director of Pfizer's US regenerative-medicine unit. Still, he remains bullish about the field. "With risk comes reward," he says.

And for every dead end, there's an opportunity. In April, Neusentis announced that in conjunction with Pfizer Canada, it would fund six Can\$50,000 (US\$52,000) postdoctoral fellowships this year and next with the aim of fostering high-calibre stem-cell scientists and developing new areas of clinical interest.

Ruth McKernan, Neusentis's chief scientific officer, encourages budding stem-cell scientists — among them her 20-year-old daughter — to consider jobs in industry, especially if they have clinical interests. "In the pharmaceutical industry, you can see how your unique piece of science enables the progress of drugs from research all the way to the clinic," she says. "And it's very difficult to see that spectrum in academia." **M.W.**

B. DAMONTE/RAFAEL VIÑOLY ARCHITECTS

This CIRM-funded stem-cell centre at the University of California, San Francisco, opened in February.

effort launched in 2002 with the lofty goal of becoming Europe's flagship stem-cell research centre — has roughly doubled in size, to about 250 researchers in 25 groups. And "we are actively recruiting at all levels and still growing the initiative", says Roger Pedersen, a human-ES-cell scientist and co-convener of the initiative. Meanwhile, the UK government's Technology Strategy Board has announced plans to open a cell-therapy technology and innovation centre, aimed at spurring commercialization of therapies in the United Kingdom. The board will give it between £20 million and £40 million (US$33 million to $65 million) in core funding over the next four years, according to Nick Sheppard, a spokesman for the board, and is expected to raise twice as much again from other sources, both public and private.

Some Britons might actually have an easier time in the United States. That has been the case for Paul Burridge, a British postdoc at the Johns Hopkins Institute for Cell Engineering in Baltimore, Maryland, who is working on deriving heart cells to treat cardiac disease. Drawn, he says, by the ease of obtaining funding from the Maryland Stem Cell Research Fund, and discouraged by difficulties in the United Kingdom, Burridge relocated in 2008 and plans to focus on the less controversial induced pluripotent stem (iPS) cells, which are created not from embryos, but by reprogramming adult cells. "The policy changes that have been made [in the United States] have had a drastic effect on my career," he says.

A DIVERSIFIED PORTFOLIO

It is a high-risk strategy to focus research solely on ES cells. Fortunately, diversifying has become easier: the advent of iPS cells has generated opportunities for stem-cell scientists that are much less constrained by politics. Researchers hope that human iPS cells, which behave similarly to ES cells, will help them to address basic questions about how undifferentiated cells become heart or muscle cells, as models for studying disease and as targets for drug development. Other options are provided by adult stem cells obtained from various organs. These cells are much less flexible, but come without ethical conundrums. In addition, working with non-human ES cells offers a less-politicized route to studying subjects such as developmental biology. "The field has been opened up and there are many different directions it can be taken," says Timothy Kamp, director of the University of Wisconsin Stem Cell & Regenerative Medicine Center in Madison. So the legal and political uncertainty around ES cells shouldn't "kill people's career decisions", he says.

The interest in iPS cells, in particular, isn't confined to academia. "Across the industry, this is an area where there is likely to be increased opportunity for young people," says Ruth McKernan, chief scientific officer for Neusentis, a subsidiary of Pfizer based in Cambridge, UK, that conducts the company's stem-cell and regenerative medicine work. The same techniques and science underpin iPS cell biology, so experience working with human ES cells is of increasing value to the pharmaceutical industry, says McKernan.

GET BROAD TRAINING

The basic intellectual tool kit for scientists entering the field includes a good grounding in cell and molecular biology and some understanding of developmental biology — typical components of undergraduate and graduate programmes. Kamp adds that researchers can make themselves more marketable by honing expertise in techniques beyond cell culture, such as advanced imaging technologies, or metabolomics, the systematic study of all the metabolites present in a cell, tissue, organ or organism.

Although these subjects can be studied in many departments of biology, several institutes now have more targeted programmes. In addition to Stanford's nascent stem-cell PhD, the Sackler Institute of Graduate Biomedical Sciences at New York University School of Medicine is now offering its PhD students a track specializing in stem-cell biology. The Sackler programme has been helped greatly by a $1.9-million training grant from the state's stem-cell initiative, says Joel Oppenheim, the senior associate dean for biomedical sciences at the school of medicine. His students, he notes, have not been deterred by the uncertainty around NIH funding. For them, stem-cell research "is sort of the new hot buzz topic", he says.

The University of Minnesota in Minneapolis now offers a master's degree in stem-cell biology. And Dan Kaufman, associate director of the university's stem-cell institute, says that he hopes it will become a fully fledged PhD in stem-cell biology.

Postgraduate programmes are also taking root outside of the United States. In Germany, the Hanover Biomedical Research School in Germany launched a three-year PhD in Regenerative Sciences in 2007 that accepts up to 20 students a year from all over the world. In Britain, The Wellcome Trust Centre for Stem Cell Research and the Medical Research Council together fund a highly competitive four-year doctoral programme at the University of Cambridge, and the University of Sheffield offers a master's degree in stem-cell biology.

Other programmes may provide young researchers with outside-the-box opportunities as the stem-cell field matures. Consider, for instance, the emerging need for biomanufacturers with stem-cell expertise, as exemplified by a new PhD programme in stem-cell biomanufacturing at the Georgia Institute of Technology, funded by the US National Science Foundation. The programme opened its doors last year and is admitting six students per year. "If stem cells are going to move out of the lab, there will be lots of need for engineers to produce a large number of identical cells," says Levine.

Wherever an aspiring scientist's inclinations point, there should to be plenty of activity and opportunity, politics notwithstanding. And there is certainly plenty of interest. In the first week after Stanford announced its new PhD — and before she had posted any information about it on the Internet — Reijo Pera received 82 e-mails expressing interest in applying. "This is actually a statement that stem-cell biology and regenerative medicine are here to stay," she says. "It is a discipline with unique skills and training. It is a real entity, not hype." ■

Meredith Wadman *is a* Nature *correspondent based in Washington DC.*

NON-SKID

A sticky problem.

BY JOHN FRIZELL

Ellie stared aghast at the mirror. There was a bruise the size of a cantaloupe on her thigh and its colours seemed to be getting more livid as she watched. It hurt but the physical pain didn't come close to matching the mental anguish. There was a pool party tomorrow and she couldn't think of wearing a bikini. Even a bare-midriff look was out of the question. She could only wear about half of her party clothes, and she had a lot of parties to attend over the Christmas season. It wasn't fair. And Kathi had broken her wrist. Something had to be done.

She pulled her skirt back down, gingerly patted everything into a clean line and then phoned Jamie.

"Why don't you come down to the kitchen? I'm going to make some of Mom's eggnog."

"Great. You can tell me what you want when I get there."

Jamie was only a few metres away but of course he was in his room — where else? — and Ellie had decided never to go there again after having been trapped by one of his robots.

"There is ice all over everything outside," she said as she grated nutmeg and measured out chopped vanilla pod.

"Common this time of year."

"I'm falling over. Getting bruised."

"Don't go out. We have optical broadband. You can get everything you need."

"Jamie!"

"Walk carefully."

"I do. It's not enough."

"Wear crampons. You can order instep crampons off the Internet. Don't go out until they arrive."

He took two forks out of the drawer and showed her how the little metal claws could be attached to her shoes. Great. She would be walking around like some sort of predatory animal, ruining her shoes, with a big ugly strap running over the top of them to hold the device in place. Ugh. No way.

"That's a really good idea Jamie. But they might not be quite the right look."

Jamie's face went vacant, as it always did when she talked about any aspect of style.

"But I bet you could make something that would do the same job but be invisible. Of course no one else has managed to …"

She watched as his face became animated again. It was a bit unfair to manipulate him like this, but her brother loved technical challenges and it would be fun for him.

"Eggnogs for a week," he said.

"You put it on like this," Jamie said, brushing a thin brown goo onto the soles of her oldest and worst shoes. "Don't get it on your hands."

He was wearing disposable gloves.

"What happens if I get it on my hands?"

"Just don't."

"But suppose I touch the soles of my shoes while I am putting them on?"

"No problem. Once this stuff has been in contact with the soles for 3 or 4 seconds it beds irreversibly into the material. It prefers polyurethane or PVC but it will adapt to whatever it finds. It works by …"

Ellie forced herself to listen, an intent expression on her face, nodding or saying 'Oh, really' when he paused, but as always it went right over her head. There were nano machines and long chains with ions on them or something, but the more Jamie explained the less she understood.

"Can I try them?"

She took a few steps towards the front door.

"They feel the same as ever."

"The nanos recognize flooring materials and inactivate. Go outside."

She put on a warm coat and gingerly stepped onto the icy front path. It felt fine. She took hold of the fence just to be safe and lifted a foot. She could balance. She let go of the fence. She could still balance. She walked up and down the slick gleaming ice of the path in perfect comfort.

"I got the specific adhesion perfect didn't I," said Jamie.

"You are a genius," said Ellie, kissing him on the cheek. She thought of her many friends and of Kathi, stuck inside with her arm in a cast. "Can you make a bit more of this?"

Ellie's social standing, already much better than average, had gone through the roof since she had started treating shoes. At first people brought old beat-up shoes and then came back a day or so later, but as her reputation spread people started showing up with their best. Girls who would not normally mix with anyone in her set were suddenly including her in their circle. She was careful. She never let anyone else touch Jamie's stuff and she always wore gloves — she had gone through two boxes of disposable latex gloves. She didn't really like the nickname 'Shoe Queen', but the 'Queen' part was good.

➲ NATURE.COM
Follow Futures on Facebook at:
go.nature.com/mtoodm

The icy grip of winter wasn't as bad when you could get around easily, and she had lots of opportunity to use her upgraded footwear attending parties she never would have been invited to before. It wasn't until late February that the problems started.

"Jamie!"

"Sis."

"I was walking home just now and I stepped on a patch of grass."

She waved her second best boot in his face. The sole was carpeted with grass. Some of it had roots on it with bits of earth stuck in them. She plucked at it, bits of root came off but the grass stayed as if it were part of the boot.

"I could barely get it loose. If I'd stepped on the grass with both feet I'd probably still be there."

Jamie hesitated for a moment.

"Grass isn't a floor material. The nanos don't deactivate."

"But it's thawing. In a week or so the ice will all be gone. What's going to happen then?"

"Good point. I didn't think of that." ■

John Frizell *was trained in biochemistry and works on ocean conservation for Greenpeace. In his spare time he walks, builds robots and writes short stories.*

nature OUTLOOK

HEPATITIS C

9 June 2011 / Vol 474 / Issue No. 7350

COVER ART: NIK SPENCER

The US Food and Drug Administration's approval of two new medications for hepatitis C in May marked the dawn of an era in treating this insidious disease.

The hepatitis C virus (HCV) can lurk silently for decades. Indeed, most of those who will benefit from this advance have no idea about their improved prospects, because they are unaware they are infected. The coming years could bring a surge in cases of the chronic form in people who unwittingly contracted the virus back in the 1960s and 70s through sharing needles and blood transfusions. Many of these chronic infections will progress to cirrhosis of the liver, and some to liver cancer and liver failure. Hepatitis C is the leading reason for liver transplantations. Until now, the only treatment option has been a gruelling year-long regimen of interferon-α plus ribavirin. Cruelly, in people with a particular genetic make-up, even this nasty course of medication often doesn't work.

The newly approved antiviral drugs lead the way for many other therapies (page S5). But caution is in order. First, the new medications supplement — not replace — the current treatment. So patients will be subject to all the same side effects as before, plus possibly a few more. Moreover, as clinician and HCV activist Diana Sylvestre points out on page S11, there is a high risk that the virus will become resistant to the new drugs. Finally, on page S18 epidemiologist Brian Edlin argues US policy makers need to wake up to the lurking HCV threat, and take action. Meanwhile, Egypt is experiencing an HCV epidemic of greater intensity than that in the United States and Europe (page S12), exacerbated by the fact that the genotype of HCV in Egypt is rarely found elsewhere.

This *Outlook* was made possible by the financial support of Gilead Sciences, Inc., Merck and Boehringer Ingelheim. As always, *Nature* retains full responsibility for all editorial content.

Herb Brody
Supplements Editor, Nature Outlook.

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the Nature Outlook Editorial guidelines available at http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf

CITING THE OUTLOOK
Cite as a supplement to *Nature*, for example, Nature **Vol XXX**, No. XXXX Suppl, Sxx–Sxx (2011). To cite previously published articles from the collection, please use the original citation, which can be found at the start of each article.

VISIT THE OUTLOOK ONLINE
The *Nature Outlook Hepatitis C* supplement can be found at http://www.nature.com/nature/outlook/hepatitis-c/

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES
For UK/Europe (excluding Japan):Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea:Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES
Feedback@nature.com

CONTENTS

COLLECTION

INTRODUCTION

A smouldering public-health crisis

Long overshadowed by HIV, the hepatitis C virus is starting to take its toll. And the heat is on to find and treat those affected.

BY LAUREN GRAVITZ

In the early 1980s, when AIDS was still unexplained and HIV was spreading unchecked, another blood-borne virus was also on the move. Like HIV, this unknown infectious agent struck intravenous drug users and blood transfusion recipients. But it was stealthier than HIV, often causing its victims no discomfort as it multiplied undetected for many years, even decades. The virus manifested itself slowly, starting with flu-like symptoms, such as fever and fatigue, and gradually working up to a fully fledged attack on the liver, where it could cause cirrhosis or cancer[1].

That culprit, identified in 1989, was hepatitis C virus (HCV). Some 130–200 million people are now estimated to be infected worldwide. Rates of transmission in the United States, Europe and Japan have plummeted since the virus was identified, however, thanks to disposable medical instruments and a screened blood supply. But the virus continues to thrive in developing nations, which lack the resources to treat people who do not appear ill. Treatment is expensive, lengthy and causes numerous side effects — and, for all that, it works only about half the time. The upshot is that more than 350,000 people worldwide die from HCV-related liver disease every year[2].

Hepatitis C is just as deadly as HIV — both kill about 10,000–15,000 people per year in the United States, according to David Thomas, an infectious diseases specialist at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. But HIV gets about 30 times more research funding in the United States than HCV, he says. And as increasing numbers of chronic hepatitis C cases manifest their most serious symptoms, the smouldering problem is turning into a burning public-health issue that threatens to stretch researchers, patients, and the health-care system to their limits.

THE QUIET EPIDEMIC

Two other viruses that target the liver were discovered before HCV: hepatitis A, which causes an acute infection with symptoms that fade within months, and hepatitis B, which commonly becomes chronic in children and in up to 10% of adults. HCV remained undetected for far longer because it replicates slowly and causes symptoms similar to other diseases. It was only when researchers at the National Institutes of Health (NIH) in Bethesda, Maryland, began to tease apart the causes of liver disease in the 1970s that they discovered that another infectious agent was involved.

HCV is transmitted through the blood, mainly via transfusions, shared needles and reused medical supplies. Sexual and mother-to-child transmission is much less likely than for HIV. In developed nations, most new infections occur in injection-drug users. An estimated 1.6% of the US population carries HCV, as many as three-quarters of whom don't know they have it[3]. In poorer countries, the reuse of medical supplies is still common and — in combination with a lack of screening of blood donations — is fuelling the virus's spread. Complicating the picture, HCV exists in at least eleven variations or at lest six genotypes, and treatment success varies by genotype. (see 'Spread of HCV').

> ***It's turning into a burning public-health issue that threatens to stretch researchers, patients, and the healthcare system to their limits.***

Different countries confront distinct challenges when dealing with the HCV epidemic. The Egyptian healthcare system, for instance, has to cope with the highest rate of transmission in the world, at least 14% of Egypt's citizens infected — three times the global infection rate (see 'A uniquely Egyptian epidemic', page S12). Elsewhere, severity of the HCV epidemic is obscured by other, more immediately severe public health concerns. Throughout much of Africa, HCV 'hides behind' widespread HIV and hepatitis B virus infections, says virologist Jean-Michel Pawlotsky, director of the French National Reference Center for Hepatitis B, C and delta. On a recent visit to a country in southern Africa, Pawlotsky noticed that they screen blood for HIV and hepatitis B but that "they don't have the resources" to screen for HCV.

"All countries share a few common challenges with hepatitis C," says John Ward, director of the viral hepatitis program at the US Centers for Disease Control and Prevention. "One is a very low level of awareness of the severity of the problem — clinician knowledge of HCV seems to be inadequate no matter what country you're in."

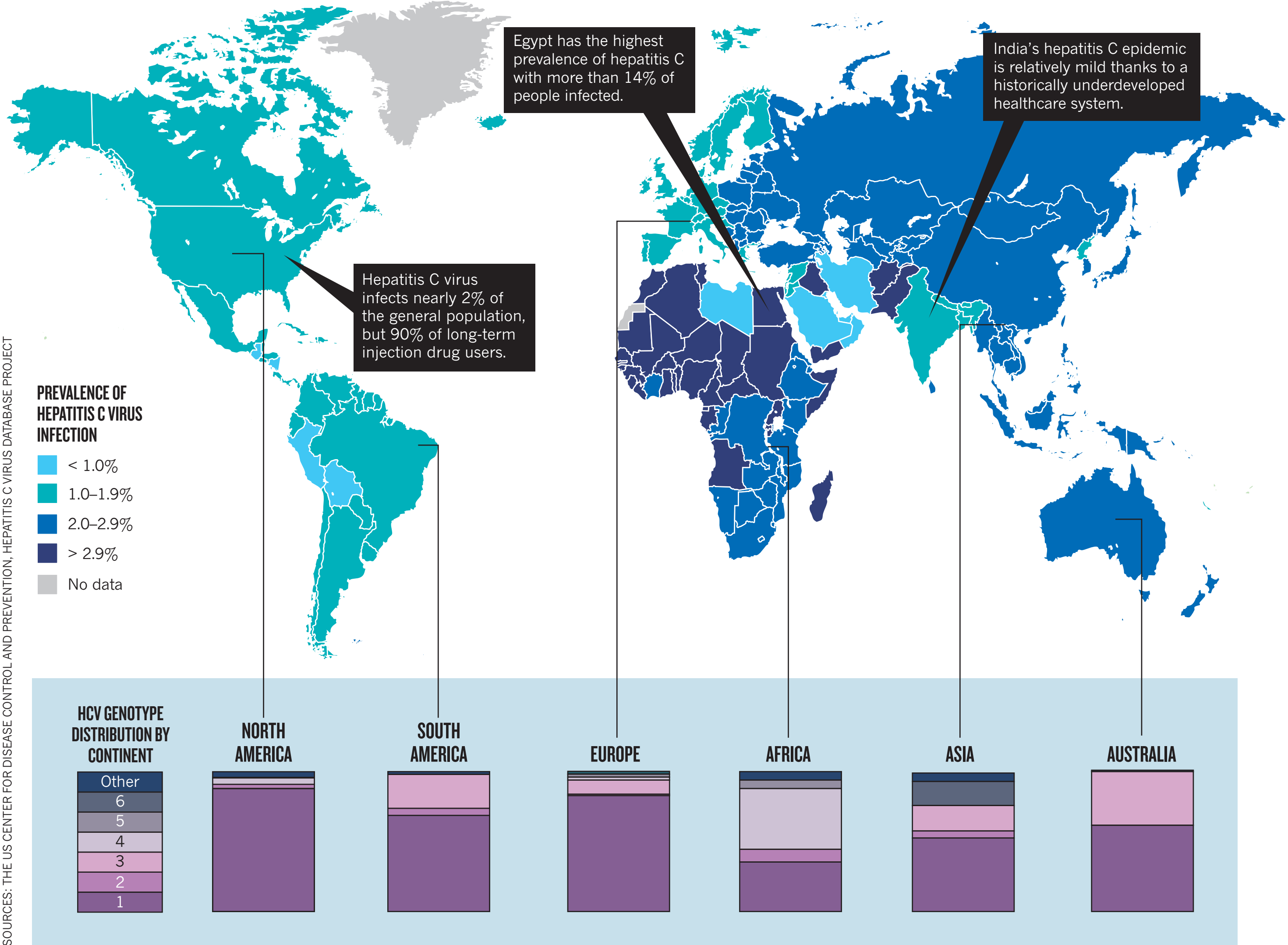
Even when a country is acutely aware of its HCV problem, the best available options are barely adequate to the task. In France, where politicians were found guilty of distributing HIV-contaminated blood in the mid-1980s, officials are now trying to prevent a repeat with HCV. French public-health workers have been screening the populace intensively for HCV, and the French government estimates that it has identified 60% of the country's infected citizens, Pawlotsky says. "Half of those have been treated, and half of those treated have been cured." Which, he says, means that "with the heaviest campaign worldwide and the largest proportion of screened patients, we have only cured twelve to fifteen percent" of those infected nationwide.

Pawlotsky and others note that there has been progress towards preventing HCV infection. This success has come in part because the rise in HCV infections coincided with that of HIV, and the HIV-related public-health campaign targeted practices (such as sharing needles) that also contributed to HCV transmission. Combined with introducing thorough screening of blood donations, health officials in the United States were able to lower the infection rate from as high institution listed, as several hundred thousand per year to about 20,000 per year now. Still, "20,000 a year is not a small number," says Brian Edlin, an epidemiologist and infectious diseases specialist at the State University of New York Downstate College of Medicine in Brooklyn, New York. "It's still a fire that's spreading — not as quickly as in the past, but it's spreading."

According to Ward, HCV infections are on the rise in older teenagers and young adults. Of those already infected, the groups most at risk of becoming symptomatic in the United States are veterans, who have an infection rate at least three times that of the general population, and

THE SPREAD OF HCV

The hepatitis C virus reaches across the globe with highest prevalence in north Africa and south Asia. A major challenge is tailoring treatments and vaccines to the various viral genotypes which affect treatment response.



SOURCES: THE US CENTER FOR DISEASE CONTROL AND PREVENTION, HEPATITIS C VIRUS DATABASE PROJECT

baby boomers (those born between 1946 and 1964). In fact, baby boomers, who make up about 30% of the US population, account for two-thirds of the people in the United States with HCV.

The potential healthcare burden these numbers represent is alarming. Over the past five years, the number of veterans with HCV who have liver cancer has tripled[4]. Some Veterans Affairs hospitals now report more cases of liver cancer than colon cancer, says Janet Durfee, who directs nationwide public-health programmes for the Department of Veterans Affairs. By contrast, in the general population, colon cancer is about four times more prevalent.

HOPE FOR NEW DRUGS

Much about HCV remains to be fully understood. It remains a mystery, for instance, why some people inflected with HCV never develop symptoms. About 15–20% of those infected with HCV can clear the virus without pharmaceutical help, and among teenagers and young adults, the proportion can climb as high as 50% (ref. 5). Of those who can't fend off the virus on their own, most remain stable for decades, without any major symptoms. In the 10–15% of people who have symptoms, however, the virus causes cirrhosis: the liver becomes fibrotic and scarred, resulting in jaundice and a swollen abdomen, as well as a dangerous build-up of toxins in the blood and other serious complications. In some patients with cirrhosis, the disease advances slowly, but one in four — 2–4% of all HCV cases — develops liver cancer or liver failure.

"The majority of people will not go on to develop cirrhosis if they don't have other viruses or co-morbidities such as alcoholism," says Harvey Alter, an NIH virologist whose research led to the discovery of HCV. "Seventy to eighty percent do well without treatment. But we can only say that for the first 30 years or so, because that's all we have data for."

The standard treatment for HCV infection is two daily doses of ribavirin, a nonspecific antiviral agent, combined with a weekly injection of interferon-α, which activates the immune system and impedes viral replication. Not only is the treatment prolonged, lasting either 24 weeks or 48 weeks depending on the virus genotype, but it can cause serious side effects, ranging from fatigue and flu-like symptoms to anaemia and severe depression. Patients, and their healthcare providers, would welcome anything that might shorten treatment, reduce its side effects and increase its efficacy. "Getting away from interferon is everbody's goal," says Alan Perelson, a mathematical and theoretical biologist who specializes in HCV at Los Alamos National Laboratory in New Mexico. "It comes with side effects, cost,

inconvenience, injections, and the fact that not everybody responds to it."

Changes are afoot. Two new antivirals are speeding their way to market. Boceprevir, made by drug giant Merck, headquartered in Whitehouse Station, New Jersey, and telaprevir, made by Vertex Pharmaceuticals, based in Cambridge, Massachusetts, are both inhibitors of an important viral protein, the NS3/4A protease. The drugs are designed specifically to attack HCV genotype 1, which is the most prevalent genotype, accounting for about 60% of global infections, and the least responsive to current treatment. They will supplement, but not replace, the standard interferon plus ribavirin therapy. Thus, they won't eliminate the cocktail's difficult side effects and may introduce new ones. The good news is that in phase III trials, each drug, in combination with the standard treatment, increased the cure rate of people with HCV genotype 1 from under half to 70%, while cutting treatment time in half for some.

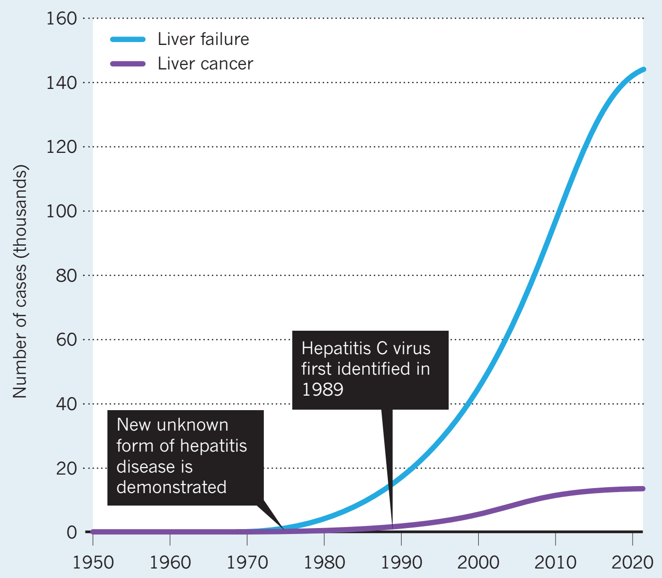
"Just one new antiviral added to the standard of care will make a huge difference," says Charles Rice, executive and scientific director of the Center for the Study of Hepatitis C in New York. "But the game's not over. We'd really like to eliminate the current standard of care and replace it with something that works better and has fewer side effects." There are several promising therapies at various points in the pharmaceutical pipeline (see 'New drugs hit the target', page S5). But there is a built-in delay in development: every new drug must be tested against the current, year-long, treatment regimen.

And drug developers are facing an elusive foe. The virus replicates so rapidly and is so error-prone when it does, that many genetic variations — including those that confer drug resistance — exist inside someone before he or she is diagnosed. The best strategy is therefore to throw multiple drugs at the virus at once, enough such that the pathogen would be unable to mutate to resist them all. Statisticians calculate such a feat would take two or three antivirals, all of which must target different components of the virus (a strategy similar to the drug cocktails used to fight HIV). The number of drugs needed "will depend on how hard it is for the virus to generate the number of mutations required to gain resistance," Perelson says.

"We're looking at an epidemic that is treatable and ultimately curable. And yet, as a nation, we've got our heads in the sand."

Further complicating the problem, because the risk factors for infection with HCV and HIV are similar, many people who are HIV-infected are also infected with HCV (up to 30% in the United States). For reasons that are not yet understood, this co-infected group has a lower cure rate and suffers from more side effects. Moreover, many of the antiretroviral drugs used against HIV can build up in the liver, poisoning the organ that the HCV drugs are working to save. And the new protease-inhibitor drugs have not yet been vetted in more than a few dozen co-infected patients. "The amount of information we have on those who are dually infected is abysmal," says Thomas. "We're going to have these treatments available this year and not have any idea on how to use them in combination with antiretrovirals."

THE COMING PROBLEM
The projected medical impact of the hepatitis C virus in the United States.



SOURCE: DAVIS GL ET AL. GASTROENTEROLOGY 138, 513–521 (2010).

Most researchers and public health experts concur that, in addition to better treatments, the ultimate solution is a hepatitis C vaccine. "Otherwise, we'll have infection from other parts of the world where therapies are not available," Pawlotsky says. But vaccine development is challenging, and the virus has kept researchers guessing for decades already (see 'A moving target', page S16).

OUT OF HIV'S SHADOW

When discussing HCV, comparisons with HIV are nearly unavoidable — both appeared on the communicable disease radar around the same time and both have similar modes of transmission. But the HCV problem cannot be solved by repurposing the agents developed to fight HIV, given the marked differences in the biology of these viruses. "Their differences are greater than their similarities," says Miriam Alter, an epidemiologist at the University of Texas Medical Branch in Galveston. She notes that HCV has been endemic to human populations for centuries, and it is biologically much different from HIV.

The differences also stretch beyond biology and into public-health initiatives. The discovery of HIV led to a huge amount of funding for antiviral research and development around the world. "With hepatitis C," says Thomas, "we haven't even had widespread adoption of treatment here in the US."

Even diagnosis of HCV infection (see 'A testing journey', page S20) has lagged far behind HIV. The CDC estimates that about 80% of HIV-infected people in the United States know they have the disease, but the figure for HCV awareness is about 30%. And worldwide, that number shrinks to an abysmal 5%. "We're looking at an epidemic that is five times the size of the HIV epidemic, that spreads more quickly, that is treatable and ultimately curable. And yet, as a nation, we've got our heads in the sand," says Edlin.

At the same time, it's hard to deny that the field has made great strides in a short time. As Rice says: "In 1975, there was this agent causing all this disease, and nobody knew what it was. And here we are today and we can effectively treat a majority of people who are infected." The research is headed in the right direction — but disease awareness still lags far behind. It will take a concerted effort and a huge national and international collaboration among public-health officials, doctors and governments to stop the raging epidemic before it burns through the limited resources of healthcare systems around the world. ■

Lauren Gravitz *is a writer in Los Angeles, California.*

1. Institute of Medicine. Hepatitis and Liver Cancer: A National Strategy for Prevention and Control of Hepatitis B and C (National Academies Press, 2010).
2. Perz, J. F., Armstrong, G. L., Farrington, L. A., Hutin, Y. J. F. & Bell, B. P. *J. Hepatol.* **45**, 529–538 (2006).
3. Alter, MJ. Epidemiology and burden of hepatitis C in the United States, in Hepatitis C and Liver Transplantation, (ed. by Sandeep Mukherjee, 2011).
4. Department of Veterans Affairs, Veterans Health Administration, Office of Public Health and Environmental Hazards and Public Health Strategic Health Care Group State of Care for Veterans with Chronic Hepatitis C (Department of Veterans Affairs, 2010).
5. Centers for Disease Control and Prevention. *MMWR Recomm. Rec.* **47** (RR-19), 1–39 (1998).

MARK VIKER/GETTY IMAGES.

THERAPEUTICS

New drugs hit the target

With two recently approved drugs and dozens more in the pipeline, hepatitis C treatment will improve over the next decade.

BY JANA SCHLÜTTER

When Charles Gore talks about some of his colleagues, there is more than a hint of urgency in his voice. Although he cleared his hepatitis C virus (HCV) infection after receiving the standard treatment, two of his staff at the World Hepatitis Alliance, an advocacy organization, recently had liver transplants. "And they are lucky," says Gore, who is president of the alliance. "This treatment does not help about 50% of the patients who are infected with the most common form of the virus. So their liver becomes worse, and many of them cannot get a transplant. They are facing death."

Around the globe, patients who have not been cured by the current treatment, a combination of interferon-α and ribavirin, are waiting for new drugs. So far, their doctors have had nothing to offer them but another 48-week-round of the same drug combination, which had its last upgrade in 2001 when researchers attached a molecule called polyethylene glycol to interferon-α. This 'pegylation' allows interferon-α to stay in the body much longer, reducing the frequency of injections from three per week to one. But the side effects are just as harsh, including flu-like symptoms, anaemia and depression. And although the patient being treated may be too weak to work or enjoy family life, the virus often manages to survive and prosper under these conditions. At most, 20% of patients are cured by this second course of treatment. Still, there was no alternative.

This situation is about to change. Two powerful weapons against chronic HCV infection have been licensed: the protease inhibitors telaprevir, from Vertex Pharmaceuticals, based in Cambridge, Massachusetts, and boceprevir, from drug company Merck, headquartered in Whitehouse Station, New Jersey. When either drug is added to the current therapy, the cure rate increases for patients who have so far been spared the daunting year-long treatment: that is, 'treatment-naive' patients. The drugs also offer hope to those increasingly desperate patients who have not been helped by the standard treatment: instead of around a 20% chance of a cure, these 'treatment-experienced' patients now have a 30–90% chance. "We are approaching a new era of management of this disease," says Mark Thursz, a hepatologist at Imperial College London and current secretary-general of the European Association for the Study of the Liver (EASL).

➲ NATURE.COM
to read the latest news on hepatitis drugs and treatments
go.nature.com/RPNcq0

The drug manufacturers have tailored these protease inhibitors to HCV genotype 1, one of at least six forms of HCV. Genotype 1 is particularly widespread in the United States and Europe and is one of the least responsive to the standard treatment. The clinical studies coming out now, Thursz says, "show that the new drugs can tame the pit bull terriers of the hepatitis C world: the genotype 1 viruses."

In addition to telaprevir and boceprevir, there are dozens of compounds in the pipeline, and that's only counting the ones that drug manufacturers are willing to disclose. These drugs target many aspects of the virus's life cycle — the stages it goes through in the liver cell to reproduce itself. Used in combination, the new agents might be able to target all HCV genotypes at once, while improving the cure rate and preventing drug resistance from emerging. Although most of these drug candidates are being added to the current treatment, an interferon-free regimen has recently shown promise — a possibility that could substantially reduce treatment side effects and increase adherence.

DIRECT HITS

In the current regimen, interferon-α boosts the patient's immune system, and ribavirin is a general inhibitor of virus replication. By contrast, the new drugs target HCV directly. Telaprevir and boceprevir block HCV's NS3/4A protease. After an HCV particle attaches to and enters a liver cell, it releases its RNA, which is subsequently translated into a single polyprotein (see 'The life of HCV'). This long chain is cleaved into functional proteins by NS3/4A, which acts like a pair of molecular scissors. Without the protease, functional viral enzymes and structural proteins are not generated, so HCV cannot complete its life cycle.

This March, the drug companies reported results of phase III clinical trials of telaprevir and boceprevir, each coupled with the current therapy, at EASL's International Liver Congress in Berlin. Two-thirds to three-quarters of treatment-naive patients with HCV genotype 1 are likely to clear the virus permanently. And treatment time is expected to be halved for patients in this group who have undetectable levels of virus after four weeks of treatment.

More hotly anticipated were the data for the treatment-experienced patients, including relapsers, whose virus had become undetectable but rebounded after their previous treatment ended; partial responders, whose

viral load decreased by at least 99% but never became undetectable; and null responders, who previously had little success in fighting the virus. Telaprevir was tested in the Realize trial, which involved 662 patients from Europe and the United States. Adding telaprevir for 12 weeks to a 48-week-treatment course increased the cure rates from 24% to as high as 88% in relapsers, from 15% to 59% in partial responders, and from 5% to 33% in null responders. Boceprevir was tested in 403 patients in centres across the United States and Europe in the Respond-2 trial. Adding boceprevir for 32–44 weeks caused the cure rate to climb from 29% to 69-75% in relapsers and from 7% to 40-52% in partial responders. (Null-responders did not participate in this trial.)

"To have direct-acting antivirals against hepatitis C and to see such increases in cure rates is a huge step forward," says Stefan Zeuzem, a hepatologist at the Goethe University Medical Center in Frankfurt, Germany, who was involved in both the Realize and Respond-2 trials. But these drugs are not cheap. "Cost will be a major issue," he says. "However, we are preventing liver cancer and other end-stage liver diseases, which makes it worthwhile. We are aiming for a cure, not just a few more weeks to live."

Both of these drugs also have side effects. More than half of the patients treated with telaprevir developed a rash, with 3–6% having a rash severe enough to halt treatment. Boceprevir is associated with anaemia (similarly to telaprevir) and can cause a metallic taste in the mouth, both of which affect nearly half of all patients. These problems are in addition to those caused by interferon-α and ribavirin, meaning that nearly every patient in the clinical trials suffered from at least one side effect. "It's still a tough treatment," says Gore. "For patients, it's very important that clinicians manage these side effects well."

If side effects cause patients to abandon treatment on the new regimen, this could lead to HCV developing resistance to the new drugs. The new protease inhibitors cannot be given alone and must be given with interferon-α and ribavirin to prevent protease-inhibitor resistance emerging. Thursz adds that as boceprevir and telaprevir are similar compounds, resistance to one will probably translate into resistance to the other (so-called cross-resistance), restricting future treatment options.

"The new drugs can tame the pit bull terriers of the hepatitis C world: the genotype 1 viruses."

HCV is a highly mutation-prone virus, with many genetic variants present in any one host. Before treatment starts, variants that are resistant to a particular drug make up a minority of the viral population. Under selective pressure of the antivirals, however, these variants could become the dominant strains. "We understand resistance and have to manage it," says Jean-Michel Pawlotsky, a hepatitis specialist at the University of East Paris in Créteil, France, and director of the French National Reference Center for Viral Hepatitis B, C and delta. He recommends that these new drugs should be administered at expert centres that can monitor resistance issues: "It is better to be well-treated than just treated," he says.

Despite the high cure rates, not every HCV-infected patient will benefit from the new drugs. Possible drug–drug interactions are not yet fully understood. And there are no data for the many patients who are co-infected with HIV or for patients with end-stage renal disease, decompensated (or extremely advanced) liver cirrhosis or a recent liver transplant. Furthermore, telaprevir and boceprevir have been licensed by the US Food and Drug Administration only for treating HCV genotype 1 infection. As Pawlotsky says, "What we are seeing now is just the first step into the era of direct-acting antivirals. It will cause a real shift, but it's not a full revolution."

COVERING EVERY ANGLE

More than 50 other drugs are, however, in the research and development pipeline (see 'Drug candidates for treating HCV infection 2011'). Many of these are in new classes — that is, they target different mechanisms — and can be combined to create antiviral cocktails, limiting the emergence of drug resistance. With so many new agents snapping at their heels, boceprevir and telaprevir might have a very limited time as the dominant new drugs, says Zeuzem.

Two other first-generation protease inhibitors are in phase III trials: TMC435, from Tibotec Pharmaceuticals, in Beerse, Belgium, and pharmaceutical company Medivir in Huddinge, Sweden; and BI201335, from pharmaceutical company Boehringer Ingelheim, headquartered in Ingelheim am Rhein, Germany. Both are taken once daily instead of three times, seem to cause fewer side effects and might even be more potent than boceprevir and telaprevir.

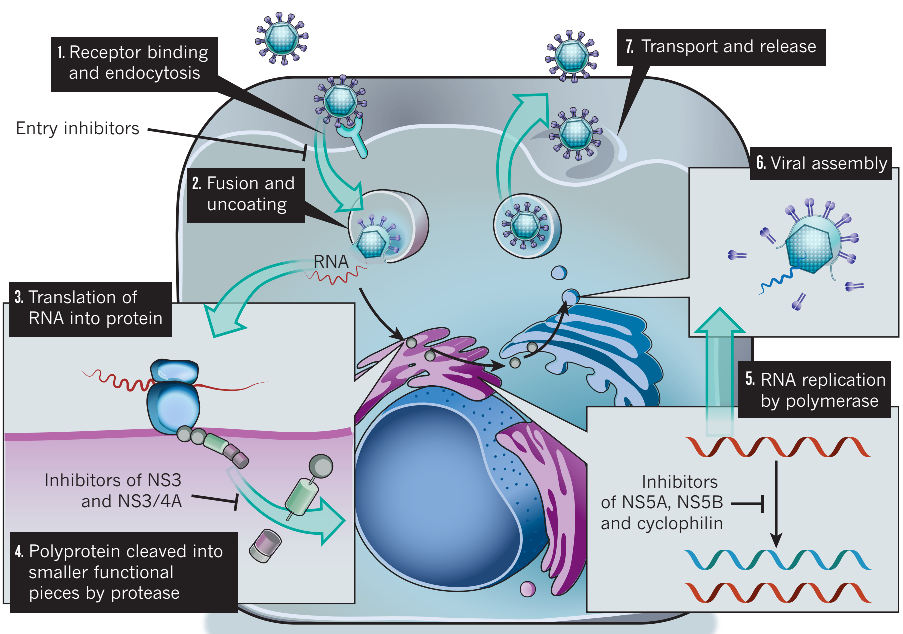
The second generation of protease inhibitors is expected to be led by Merck's MK-5172, a compound that does not seem to have cross-resistance issues with other drugs of this class and might be effective across multiple genotypes. "We want to see if the resistance profile is robust enough that we can treat people who are failures from earlier generations of protease inhibitors," says Keith Gottesdiener, vice president for hepatitis C clinical development at Merck. "That would be exciting if it was proven in the clinic."

The pharmaceutical company F. Hoffmann-La Roche, headquartered in Basel, is about to start phase III trials of mericitabine, which blocks the activity of HCV's polymerase enzyme, NS5B. By mimicking the building blocks of RNA, mericitabine is incorporated into newly formed viral RNA but prematurely terminates it, halting the life cycle.

Another protein generating immense interest as a drug target is NS5A. Its precise function is mysterious, but it seems to be involved in the replication, assembly and release of HCV.

THE LIFE OF HCV

Once it has infected liver cells, HCV goes through discrete stages of replication that are targeted by drugs in development.



SOURCE: CIESEK, S. & MANNS, M. P. NAT. REV. GASTROENTEROL. HEPATOL. 8, 69–71 (2011).

DRUG CANDIDATES FOR TREATING HCV INFECTION 2011
A sample of some of the novel agents in development to target HCV

	Direct-acting antiviral agents				Host-targeting agents			
Mechanism	**Inhibitor of polyprotein processing**	**Inhibitor of HCV replication**			**Anti-apoptotic agent**	**Antiviral agent**	**Immunomodulatory agent**	**Inhibitor of virus fusion with host cell**
Target	NS3 or NS3/NS4A protease	NS5A	NS5B polymerase		Caspases	Cyclophilins	Interferons	Viral entry
			Nucleoside analogue	Non-nucleoside inhibitor				
Recently approved	Telaprevir (Vertex) Boceprevir (Merck)	None	None	None	None	None	None	None
Phase III	TMC435 (Tibotec and Medivir) BI201335 (Boehringer Ingelheim)	None	None	None	None	Alisporivir (DEB025; Novartis)	None	None
Phase II	ACH-1625 (Achillion) BMS-650032 (Bristol-Myers Squibb) BMS-791325 (Bristol-Myers Squibb) Danoprevir (RG7227; Roche) GS-9256 (Gilead) GS-9451 (Gilead) ABT-450/r (Abbott and Enanta) Vaniprevir (MK-7009; Merck)	ABT-267 (Abbott) BMS-790052 (Bristol-Myers Squibb) GS-5885 (Gilead)	IDX184 (Idenix) Mericitabine (RG7128; Roche) PSI-7977 and PSI-7851 (Pharmasset) RG7128 (Roche and Pharmasset)	ABT-333 (Abbott) ABT-072 (Abbott) ANA598 (Anadys) BI207127 (Boehringer Ingelheim) Filibuvir (Pfizer) IDX375 (Idenix) Tegobuvir (GS-9190; Gilead) VCH-916 (Vertex) VX-222 (Vertex)	IDN-6556 (Idun/Conatus)	NIM811 (Novartis) SCY-635 (Scynexis)	PEGylated interferon-λ (Bristol-Myers Squibb)	None
Phase I	GSK2336805 (GlaxoSmithKline) IDX320 (Idenix) MK-5172 (Merck) VX-985 (Vertex)	AZD7295 (AstraZeneca) PPI-461 (Presidio)	GS-6620 (Gilead) INX-08189 (Inhibitex) PSI-938 (Pharmasset)	GSK2485852 (GlaxoSmithKline) VX-759 (VCH-759; Vertex) GS-9669 (Gilead)	None	None	GS-9620 (Gilead)	ITX-5061 (iTherX)
Preclinical	ACH-1095 (Achillion) ACH-2684 (Achillion) AVL-192 (Avila) GNS-227 (GenoScience Pharma)	ACH-2928 (Achillion) BMS-766 (Bristol-Myers Squibb) EDP-239 (Enanta) IDX380 and IDX719 (Idenix) PPI-437, PPI-668, PPI-833 and PPI-1301 (Presidio)	PSI-661 (Pharmasset)	BILB 1941 (Boehringer Ingelheim)	None	None	None	ITX4520 (iTherX) PRO 206 (discontinued; Progenics) REP 9C (REPLICor) SP-30 (Samaritan)

BMS-790052, from biopharmaceutical company Bristol-Myers Squibb, headquartered in New York, was the first inhibitor in this class and is now in phase II trials. The pipeline is rapidly filling with others.

Cyclophilin A inhibitors block a host protein that is essential for viral replication. Candidates include alisporivir (DEB025), from drug company Novartis, headquartered in Basel, Switzerland, which is in phase III trials. In theory, targeting a human protein that HCV needs will render the virus' genotype or mutation status irrelevant and make it much less likely that resistant strains of HCV will emerge.

FREE FROM INTERFERON

There is also hope for patients who are not responsive to — or cannot tolerate — the backbone of triple therapy: interferon. This April at the International Liver Congress, Anna Lok, a hepatologist at the University of Michigan in Ann Arbor, presented data from a small phase IIa study of an interferon-free regimen in null responders. The study comprised patients on double therapy consisting of two classes of direct-acting antiviral: Bristol-Myers Squibb's BMS-650032 (a protease inhibitor) and BMS-790052 (an NS5A inhibitor). These patients were compared with a cohort taking quadruple therapy, consisting of these two antivirals plus interferon-α and ribavirin. The quadruple therapy suppressed HCV in 10 out of 10 patients for at least 12 weeks after treatment, whereas the interferon-free double therapy suppressed HCV in 4 out of 11 patients, with 6 being null or partial responders.

The numbers might not seem great, but they are a start. "The potential for an interferon-free regimen is some of the most exciting news this year," says Thursz. Without interferon-α and ribavirin, the virus was expected to rebound after treatment, but this occurred in only one case. "There is still a lot of work to be done. But this was a group of very difficult-to-treat patients with excellent outcomes. Although the numbers are small, I think this is the direction we can expect to go in the future."

Indeed, this possibility has energized hepatitis C researchers. "People would have laughed at you if you suggested something like this five years ago," says Zeuzem. "Now, we know that such a therapy might be available in another five to ten years."

Many of the other drugs in the pipeline, such as the NS5B inhibitors, could also be candidates for an interferon-free regimen, says Paul Pockros, co-director of clinical research at the Scripps Translational Science Institute in La Jolla, California, who is involved in phase II studies of mericitabine. Although mericitabine is slightly less effective than the protease inhibitors, it seems to be a safe drug with a high barrier to resistance. "This one would be a good partner for a protease inhibitor," says Pockros.

With all the excitement about new drugs, one would be forgiven for thinking that interferon has had its day. But there is also development on this front. Bristol-Myers Squibb has developed a variant called pegylated interferon-λ, which is designed to be more potent and safer than interferon-α. Interferon-λ docks with different receptors that are less common than the receptors for interferon-α. This interferon circumvents the bone marrow and therefore avoids anaemia and flu-like symptoms, so it might be a good partner for direct-acting antivirals. "This would help a lot of people who cannot tolerate the current interferon," says Zeuzem, who is involved in clinical trials of this drug.

With interferon-free regimens on the horizon, the question is whether a new interferon will be needed. But there are many potential pitfalls on the way to the clinic, and HCV is a very difficult virus to target. Researchers need as many options at their disposal as possible, says Zeuzem, "just in case."■

Jana Schlütter *is a freelance journalist based in Berlin.*

BILL HACKWELL

PERSPECTIVE

Miles to go before we sleep

New drugs are generating much excitement, but a cure for all will take generations of therapies, argues **Charles Rice**.

The blood-borne hepatitis C virus (HCV) infects at least 130 million people worldwide. Over decades of chronic infection, patients are at risk of fibrosis, cirrhosis and liver cancer. Currently, HCV infection is treated with a weekly injection of pegylated interferon-α and a twice-daily weight-based dose of ribavirin. This standard of care, which has not changed for almost a decade, is considered suboptimal because of its long duration, side effects and inability to cure about half of all patients. In May this year, boceprevir, a first generation direct-acting HCV inhibitor, was approved for use by the Food and Drug Association. Patients and doctors have been waiting more than twenty years for this breakthrough, but the celebration should be tempered. Achieving a universal cure for HCV will require much more work.

The new drugs generating all the buzz are the inhibitors of the viral NS3/4A protease. Telaprevir (Vertex Pharmaceuticals, based in Cambridge, Massachusetts) and boceprevir (Merck, headquartered in Whitehouse Station, New Jersey) have improved the cure rates for both treatment-naive patients and those who had failed standard therapy. A major caveat of the protease inhibitors, however, is that they must be used in combination with the existing standard of care. This is because the virus can easily overcome a single direct-acting antiviral. HCV replicates in a prolific and highly error-prone manner, leading to the rapid emergence of point mutations that confer drug resistance. The continued need for interferon-α and ribavirin is disappointing because the new agents worsen the problematic side effects of standard therapy, especially rashes for telaprevir and anaemia for both telaprevir and boceprevir. Furthermore, host factors that affect the outcome of interferon-α plus ribavirin treatment continue to be reflected in protease-inhibitor clinical trials, for example African-American ethnicity, advanced liver fibrosis and single nucleotide polymorphisms in the region of the *IL28B* gene — although, to be sure, their adverse impact is reduced.

It is imperative that the HCV drug pipeline be kept flowing. The goal, an interferon-α and ribavirin-free regimen, will require multiple drugs with diverse modes of action. Cocktails of targeted antivirals will make it more difficult for the virus to become resistant — a strategy analogous to highly active antiretroviral therapy against HIV. Early combination trials are under way and showing promise, but it will probably be two to three years before two direct-acting antivirals are approved in combination with the standard of care, and perhaps five years before the first interferon-α and ribavirin-free drug cocktails reach the clinic.

➲ NATURE.COM
For a selection of Charles Rice's articles
go.nature.com/HPUX9n

As second- and third-generation combination regimens for HCV are unveiled, it is important to consider all those who need them. First, the new drugs will have to be tested in diverse patient groups. With at least six genotypes and dozens of subtypes, HCV is one of the most variable viruses known. Most clinical trials have been conducted in the context of genotype 1 — the most common type in Europe and North America. Each drug regimen will need to be tested at several geographical sites to encompass the most prevalent local genotypes. Patient factors must also be considered when evaluating treatment efficacy. Patients with cirrhosis, those receiving liver transplants and groups who are co-infected with HIV or hepatitis B virus are among those with the most critical needs and have traditionally been the hardest to treat. How these patients respond to next-generation drugs has not yet been properly evaluated.

People at the highest risk of the HCV are among the most marginalized groups in society, such as injection drug users, prisoners and those living in endemic regions of the developing world. In parts of Asia and Africa, up to 10% of the population is infected with HCV; in Egypt, the prevalence of anti-HCV antibodies is as high as 15%. As newer, more expensive, treatment options become available, the gap between those who can afford treatment and those who cannot will widen. Advocates for HIV/AIDS care made strides in promoting surveillance, increasing education, reducing stigma, strengthening health infrastructure and implementing harm reduction programmes. Importantly, tiered pricing models for antiretroviral drugs were negotiated, which ensure lower prices in poor countries and help promote universal coverage. The HCV community can learn from the HIV experience, including the need for combination therapy and the importance of integrating education and prevention when deploying new therapies.

A cure is possible — a statement that can't be made about many other chronic diseases. Direct-acting antivirals mark an important first step, but this is no time to lose momentum. Researchers in academia and industry need to keep working on the pipeline to ensure that all those infected with HCV can receive proper treatment, allowing this disease to fade into medical history. ■

Charles Rice *is head of the Laboratory of Virology and Infectious Disease at The Rockefeller University in New York.*
email: ricec@mail.rockefeller.edu

The author declares competing financial interests: go.nature.com/crCD7A

ILLUSTRATION BY SOPHIE CASSON/TREEINABOX.COM

PHARMACOGENOMICS

Playing the odds

Can doctors calculate a patient's chance of being cured by searching their DNA? Hepatitis C researchers are starting to make this a reality.

BY AMY MAXMEN

"Girl, you've won the lottery," said Deborah Teeters' doctor, when the results of her genetic test came back revealing two Cs at a spot among the 3 billion base pairs of her genome. Teeters, a retired child-welfare reform worker in North Carolina, had avoided treatment for hepatitis C for more than a decade because of its ugly side effects, including anaemia, fevers and severe depression. She also knew that for roughly half of all patients the 48-week regimen doesn't work. But two Cs means the odds are in her favour. She is ready to give it a shot.

Hidden within the scratch card of our genomes lie clues to how each individual uniquely responds to stress, disease and medication. If single 'letter' variations, called single nucleotide polymorphisms (SNPs), with high impact can be found, then doctors might be able to use this information to personalize patient care, and researchers could delve into the function of the affected genes to better understand disease and improve therapies. Hepatitis C now leads the movement to combine genomics and medicine, following the discovery[1] of a SNP near the gene *IL28B*. In this location, each person has either a cytidine (C) or a thymidine (T) nucleotide. And, as everyone inherits one *IL28B* gene from each parent, there are three possible combinations: CC, CT and TT. Patients with two Cs tend to clear the hepatitis C virus (HCV) when treated, whereas a CT or a TT genotype correlates with a poorer response (see 'Lucky Cs,').

A test for this SNP now helps patients decide whether to undergo treatment — which currently consists of a year-long course of interferon-α injections plus multiple daily oral doses of ribavirin — or to wait until improved drugs hit the market. And pharmaceutical companies are interested in using the test to tailor their new drugs to specific populations. In this way, hepatitis C is a success story among those who use genome-wide association studies (GWAS) to search the genome for SNPs that are clinically relevant.

"*IL28B* was a fantastic hit because nothing had ever proved as useful in GWAS before," says Ellie Barnes, a clinician scientist at the University of Oxford's Nuffield Department of Clinical Medicine, in the United Kingdom. "There have been lots of genes associated with diseases, but nothing else I know of has been put to use in the clinic and in clinical trials."

"The challenge is not only to identify biomarkers but to keep them coordinated with each compound."

It's been a whirlwind journey. *IL28B* SNPs were first linked to treatment response in late 2009 (refs 1–3) — less than a year later, doctors and pharmaceutical companies were ordering *IL28B* SNP tests. Based on the C/T SNP1, the first test was offered for about US$150 in July 2010, by LabCorp, a diagnostics company based in Burlington, North Carolina. Since then, "the test has been going gangbusters," says John McHutchison, a co-author of one of the *IL28B* papers[1] who is now at biopharmaceutical company Gilead Sciences (Foster City, California). Indeed, in April 2011, another company, Quest Diagnostics (Madison, New Jersey), launched its own version of the test.

For the majority of HCV infections in the West, the *IL28B* SNP is a more accurate predictor, or biomarker, of an individual's response to current drugs than existing markers such as viral load, ethnicity or body mass. Doctors may advise TT genotype patients to wait for better drugs as long as their liver looks healthy, McHutchison says. Or, if the side effects begin to distress a CC patient undergoing therapy, a doctor might encourage them to persist because the chance of success is high.

As more genetic factors are uncovered, not only will doctors be better placed to

recommend treatment but also researchers will learn how our immune system interacts with HCV.

STACKING THE DECK

Even before uncovering the significance of *IL28B*, there were hints that response to HCV treatment had a genetic component. Response varies predictably with a patient's ethnic group: after treatment, only 20% of African Americans successfully cleared the virus compared with 50% of people of European descent and 75% of individuals of Asian descent, explains David Goldstein, director of the Center for Human Genome Variation at Duke University Medical Center, in Durham, North Carolina, and co-author of one of the *IL28B* papers[1]. "Some people thought these variations were due to differences in behaviour, lifestyle, health care or diet, but I always suspected genes were behind it," says Goldstein. "When there's a difference this big between individuals from separate ancestries, to me the simplest explanation is a difference in genetic frequencies."

Indeed, Goldstein says, the C/T variant near *IL28B* accounts for at least half of the discrepancy between the response of African Americans and other HCV-infected patients. Whereas only about 13% of African Americans carry the lucky CC combination, about 51% of European-Americans and more than 90% of East Asians do. But other factors are clearly in play.

Two recent additions to the HCV biomarker tool kit are *ITPA* and IP-10. A GWAS at Duke found an association between anaemia in response to ribavirin treatment and two variants in *ITPA*[4]. The 'protective' *ITPA* variants reduce the activity of the enzyme inosine triphosphatase (ITPA). In the presence of ribavirin, ITPA indirectly contributes to red blood cell instability, so its disruption can protect against anaemia[5]. Predicting this outcome might be beneficial because anaemia causes up to 15% of patients to cease or reduce their treatment.

IP-10 (also called CXCL10) is a cytokine that attracts immune cells, and its levels naturally vary among individuals. Patients with high levels of IP-10 in their blood are less able to clear HCV when treated than patients with low levels. One explanation for this paradox is that some of the abundant IP-10 is in a form that blocks the signals directing other immune cells to the site[6].

Then there are other clinical clues, such as the patient's age and whether the patient has been treated before. Together, these markers help predict the likelihood that an HCV-infected patient will comply with treatment and be cured with the existing regimen.

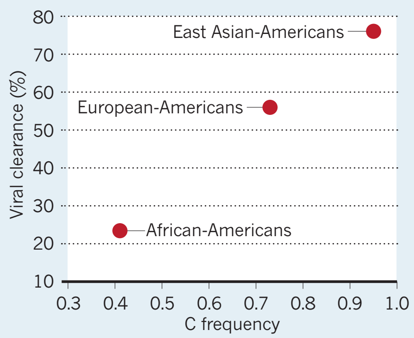
But the ground is about to shift, with the arrival of two medications: boceprevir, from the drug giant Merck (Whitehouse Station, New Jersey), and telaprevir, from Vertex Pharmaceuticals (Cambridge, Massachusetts). Trials of these drugs were in progress when the *IL28B* reports were published. Leaders at both companies say they immediately recognized the reports' significance, but it was too late to divide trial subjects by *IL28B* status. So the companies tested willing participants after the trials had begun and retrospectively analysed the data.

➲ NATURE.COM
for more on how polymorphisms influence treatment
go.nature.com/X982fs

SHAKING THE STATS

In late March 2011, Vertex and Merck presented results of their late-stage clinical trials, including data on *IL28B* status, at the International Liver Congress in Berlin. For all *IL28B* genotypes (CC, CT, TT), more patients were cured with the standard combination of interferon-α and ribavirin plus either boceprevir or telaprevir than with the standard treatment alone. In particular, TT individuals receiving treatment for the first time had much better response rates — 73% were cured by the telaprevir-containing regimen compared to about 23% with the standard regimen. But these results should be considered with caution, as the patients who agreed to be genotyped could represent a biased population.

LUCKY Cs
US citizens with two C nucleotides at a position near the IL28B gene stand a greater chance of clearing HCV when treated.



With each drug entering clinical trials, the difference between CC and TT patients seems to diminish. In an early trial of the potent antiviral agent TMC435, which is being developed by Tibotec (located in Beerse, Belgium and part of Janssen Pharmaceutical Companies) and the pharmaceutical company Medivir (Huddinge, Sweden), participants responded well almost regardless of their *IL28B* status or pre-treatment IP-10 levels. In fact, the effect of *IL28B* variants might be eliminated altogether if interferon-α were not needed, as their effects seem to depend on an interaction between proteins that originate with interferon-α and the cytokine encoded by *IL28B* (ref. 7).

But host genomics in general should help tailor treatments in the future. Certain human genotypes might predispose HCV-infected patients to developing fibrosis or liver cancer — others might be associated with severe side effects. As Jeroen Aerssens, a research fellow at Tibotec, says, "The challenge will be not only to identify biomarkers but to keep them coordinated with each compound."

FROM THE CLINIC TO THE BENCH

As doctors scramble for news on the latest biomarkers, researchers are trying to determine what these high-impact variants do. "Genome-wide association studies teach us how ignorant we are about biology," says Charles Rice, executive and scientific director of the Center for the Study of Hepatitis C in New York. "The *IL28B* association alone has catalysed a great deal of work on how the body fights infections."

The C/T variant near *IL28B* doesn't seem to alter the amount of protein produced. Instead, it must change the protein's interaction with molecules in the interferon-α pathway, the virus, immune cells or other immune-related genes. Observing immune responses and viral loads in people with different *IL28B* genotypes provides some clues. At the International Liver Congress, Zobair Younossi, hepatologist and vice president for research at Inova Health System, a large healthcare provider based in Falls Church, Virginia, reported that, before treatment, patients with the unfavourable CT and TT genotypes express more genes involved in immunity than CC patients. He speculates that this strong pre-treatment activation undermines the body's response to the drugs, which would explain why these patients are less able than CC patients to clear the virus. If he's right, temporarily suppressing specific immune responses before treatment could help those with a CT or TT genotype.

With *IL28B*, scientists have made good on a decade-old promise that genomic research will improve clinical care. Less than a year has passed since the *IL28B* test hit the market, and the biomarker is already affecting patients' lives. "I've been terrified of this treatment for ten-plus years," Deborah Teeters says, "and now I'm not as afraid, knowing I'm CC."

To researchers, it's a symbolic victory. "This is one example of how genome-wide association studies have the power to uncover genetic markers for disease propensity and treatment response," says McHutchison. "There needs to be a groundswell of activity to make people understand the potential of genomics in medicine." Hepatitis C will be at the front of the wave. ■

Amy Maxmen *is a freelance writer based in New York City.*

1. Ge, D. *et al. Nature* **461**, 399–401 (2009).
2. Tanaka, Y. *et al. Nature Genet.* **41**, 1105–1109 (2009).
3. Suppiah, V. *et al. Nature Genet.* **41**, 1100–1104 (2009).
4. Fellay, J. *et al. Nature* **464**, 405–408(2010).
5. Hitomi, Y. *et al. Gastroenterology* **140**, 1314–1321 (2011).
6. Casrouge, A. *et al. J. Clin. Invest.* **121**, 308–317 (2011).
7. Kelly, C. et al. *Gut* (online February 8, 2011).

BILL HACKWELL

Diana Sylvestre informs her patient that the hepatitis C virus has re-emerged six months after treatment.

PERSPECTIVE

Recognizing resistance

The hepatitis C virus is endemic among injection drug users, who could harbour treatment-resistant viruses. We need to adapt to this reality, says **Diana Sylvestre**.

The first antiviral agents that act directly on the hepatitis C virus (HCV) are about to hit the market. Healthcare workers have been awaiting the release of these new medications for some time, in the hope that treatment response rates would improve, even in populations of patients who are challenging to treat. But underneath the glow of anticipation lies a concern about poorly characterized risks, including the emergence of drug-resistant viral strains. The real-world impact of this risk is unclear as most of those who contract the virus do so through injection drug use and are disregarded from clinical trails.

The new protease inhibitors can elicit resistance even in patients who follow dosing regimens. But when corners are cut, risk rises. Shortening treatment, as new regimens promise to do, might reduce the burden of side effects. But the day-to-day misery will be worse with triple regimens than with the standard dual treatment, and it is important to appreciate the human tendency to reduce or skip doses of medications that make us feel ill.

Injection drug users are more complex patients: many have an unstable housing situation, unreliable transport or subject to prescription refill delays owing to insurance company bungling, which they are poorly equipped to deal with. They might be arrested and jailed during treatment. So, even though studies have shown that injection drug users have similar medication compliance rates to non-drug users[4-6], external circumstances may prevent the medication fidelity that is expected and needed.

So far, modestly reduced adherence to the interferon-α and ribavirin therapy has not led to viral resistance. Taking only 80% of the prescribed interferon and ribavirin dosages for 80% of the projected duration of treatment is sufficient to achieve optimal response rates. This allows those who treat injection drug users (including me) enough latitude to be successful. We have been able to reduce the burden of HCV in those who are most at risk of transmitting it.

Unfortunately, there is no such information on new treatment regimens. It is unclear at what point reduced adherence may become a problem. The virus rapidly mutates, so the antiviral 'pressure' exerted by the medication needs to be maintained so mutant viruses are constantly destroyed. Such protease-inhibitor-resistant strains can persist for at least three years after the withdrawal of medication[1-3]. And the conformational changes that underpin resistance to one protease inhibitor may also confer resistance to other inhibitors of that protease — a phenomenon called class resistance. And worse: if active injection drug users become reservoirs of protease inhibitor resistance, these viral strains could predominate, requiring the kind of therapeutic arms race that we see in other infections such as HIV and *Staphylococcus aureus*.

Regulators should require that clinical trials consider current or former injection drug users. This is not currently being done. The US Food and Drug Administration (FDA) Guidance for Industry document encourages trial sponsors to initiate trials early in drug development for "special populations" with unmet needs: transplant patients, people co-infected with HIV and HCV, and those with decompensated, or severe, cirrhosis. The document fails to mention injection drug users. It is as though they don't exist. If diabetics or out-of-care asthmatics were at risk, the approach would be different. Instead, the FDA has turned its back on the majority population with HCV and is approving new drugs despite having almost no understanding of their potential to cause long-term harm.

Because HCV affects those on the fringes of society, large-scale treatment studies have not been representative of the face of the disease. Their doctors are not invited to enroll them in trials. Therefore, little is known about which patients are good candidates for treatment, the importance of adherence to the treatment regimen and the outcomes in the real world — this ignorance leaves addicted HCV patients subject to the vagaries of a medical system that might not welcome them. This is unacceptable from both a humanist and a public-health standpoint.

It is time that regulators, pharmaceutical companies and healthcare workers come to terms with the fact that many patients with HCV are injection drug users. These patients must be included in safety, tolerability and efficacy trials; regulatory studies should include clinics where HCV-infected drug users are seen. And study investigators should be more representative of the kinds of doctors that usually care for these patients.

The new therapies raise the possibility of eradicating hepatitis C. But that won't happen unless the key parties in this medical drama develop a more realistic approach to understanding and treating this disease. ■

Diana Sylvestre *is an assistant clinical professor of medicine at the University of California, San Francisco.*
e-mail: diana.sylvestre@ucsf.edu

1. Forestier, N. et al. Hepatology **48** (suppl. 1), 760A (2008).
2. Tong, X. *et al. Antiviral Res.* 77, 177–185 (2008).
3. Yi, M. *et al. J. Biol. Chem.* 281, 8205–8215 (2006).
4. Tulsky JP, *et al Arch Intern Med* **160**, 697–702 (2000).
5. Singh N, *et al, Clin Infect Dis* **29**, 824-830 (1999)
6. Ferrando SJ, et al. *Am. J. Drug Alcohol Abuse* **22**, 475–487 (1996).

SHAWN BALDWIN/CORBIS

A Zabbaleen boy, from a group traditionally employed to collect rubbish, is tested for hepatitis C in Cairo.

GLOBAL HEALTH

A uniquely Egyptian epidemic

Egypt has the highest prevalence of hepatitis C worldwide. And the epidemic will soon peak. Prevention demands political will, ample funding and a change in mindset.

BY MOHAMMED YAHIA

Egypt is a hepatitis C anomaly. Globally, roughly 1 person in 50 is infected with the hepatitis C virus (HCV). In Egypt, a recent study found that about one person in seven of Egypt's 83 million population tested positive for antibodies against HCV, indicating that these individuals have been infected with the virus at some point[1]. However, nearly one person in ten carries its viral RNA and is therefore chronically infected.

This represents a huge viral reservoir that is fuelling the spread of the disease among Egyptians, mainly through hospitals and clinics (medical and dental; private and public). Firm data for the infection rate are hard to come by. A 2010 study estimated that more than half a million people are newly infected each year[2]. However, the Egyptian Ministry of Health and Population (MOHP) puts the number at 100,000 per year. Both numbers should be much lower, says Laith Abu-Raddad, an epidemiologist at the Doha-based Weill Cornell Medical College in Qatar and co-author of the 2010 study. He questions why there are so many new cases of a preventable infection each year.

Not only is the Egyptian HCV problem one of size, but the genotype of the virus in circulation is one that is not commonly found in the rest of the world. If it is not controlled, there's no guarantee that it will be confined to Egypt. Cheaper travel and increased migration means that Egypt's HCV problem is already starting to spread to Europe and the rest of the world.

SUCCESS BREEDS FAILURE

Egypt's current problem with hepatitis C arose after an earlier — successful — health campaign against a different type of pathogen. Schistosomiasis is a disease caused by a parasitic worm that can burrow through skin and lurks in contaminated water. The worm then travels to the lungs and liver, where it matures and moves to other organs. During the early part of the twentieth century, schistosomiasis was highly prevalent in Egypt, especially in rural areas, and was a major cause of liver disease. From the 1950s to the early 1980s, the MOHP and the World Health Organization (WHO) led a large-scale campaign to control the disease. Millions of people were treated with intravenous injections of tartar emetic, before an oral drug replaced this standard of care across the country in the 1980s.

Although the campaign helped reduce the prevalence of schistosomiasis, this came at a price. At the time, little was known about hepatitis or about the causative viruses and their transmission. Reuse of glass syringes and lax sterilization practices caused widespread infection with HCV, which by the 1990s had replaced schistosomiasis as the primary cause of liver disease in Egypt.

The virus that was spread by this programme has a different genotype from HCV in most other countries. Egypt — and the Middle East more generally — is a stronghold for genotype 4 (ref. 3). By contrast, about 75% of HCV infections in Europe and the United States are of genotype 1. Consequently, most research conducted in the West has gone into the control and treatment of that genotype; there have been comparatively few studies on how to improve the treatment of patients with genotype 4.

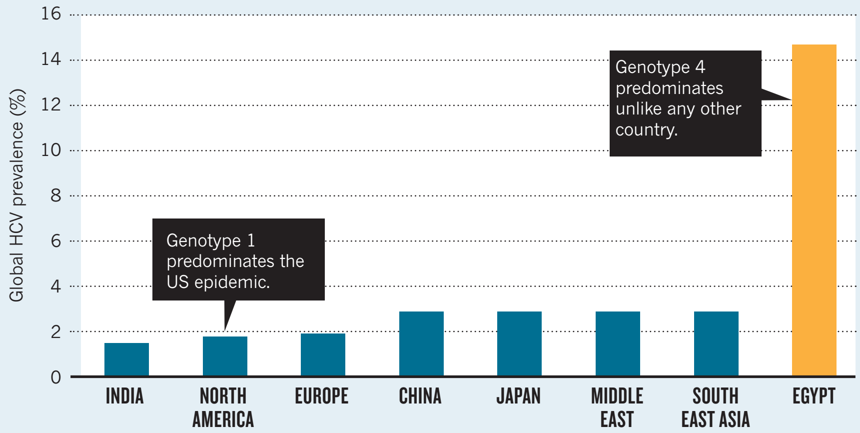
The dominance of genotype 4 in Egypt complicates the task of treating the disease. "There are no antiviral drugs being developed locally to combat genotype 4," says Alaa Ismail, head of the liver research unit at Ain Shams University in Cairo. The treatments available to people with hepatitis C in Egypt are therefore imported from the West and have been extensively tested only on genotype 1. "All the drugs that we get from overseas must undergo clinical trials here again," he says.

The best estimate for genotype 4 is that only 60% of those treated clear the virus completely. Although this is better than the cure rate for genotype 1 (at about 50%), it is still inadequate and new treatments are needed. But most Egyptians struggle to afford even the existing drugs. A full, 48-week course of HCV medication costs

NATURE.COM
visit *Nature Middle East* for latest on hep C in the region
go.nature.com/LuGXkk

EGYPT'S BURDEN

Prevalence of infection with the hepatitis C virus is nearly five times higher in Egypt than many other countries.



around US$3,500 — well above the Egyptian median income of US$2,070 per person each year.

PREVENTION IS THE BEST CURE

The MOHP has two programmes in place to tackle HCV infection — one targeted at treating chronic, symptomatic patients, and one aimed at prevention.

Through the treatment programme, the MOHP provides free or subsidized treatment for the millions of infected people who cannot afford the drugs. It is one of the largest burdens on the healthcare system in Egypt. Yet despite the large number of patients and the high cost of the treatment regimen, the MOHP argues that it is more cost effective than treating patients after they develop serious complications such as liver cirrhosis or liver cancer.

DeWolfe Miller, an epidemiologist at the University of Hawaii in Honolulu and one of the first to discover the Egyptian epidemic, in 1992(ref. 4), says: "There are national centres for HCV treatment, but there has been criticism regarding access and the quality of drugs used. I estimate that these programmes reach approximately 30% of the total number of Egyptians who are in need of treatment." Miller contends that diverting some of the MOHP's resources from treatment to prevention would be a better use of funds. "What is frequently misunderstood is that treatment, however important that may be, does not reduce or prevent transmission," he says.

At the core of the MOHP's prevention programme are public awareness campaigns, mainly run through television programmes and posters in hospitals, which have been in place for about ten years. According to Ismail, who was also director of the ministry responsible for the hepatitis C programmes — the National Hepatology and Tropical Medicine Institute (NHTMI) in Cairo — until 2006, the national centres for HCV have seen a decrease in the number of new infections over the past few years. This drop "can only be attributed to the awareness campaigns," Ismail says. "This is a socioeconomic disease, and while 60% of infections happen in hospitals and clinics, the rest are down to bad social practices, such as the sharing of razors at barber shops or the sharing of home diabetes testing equipment."

Miller, however, challenges the effectiveness of the prevention efforts. He points out that the virus continues to spread in Egypt. "It is the responsibility of all healthcare professionals to provide safe healthcare." Indeed, he says, health-care professionals could take a big step towards slowing the spread of HCV in Egypt by simply maintaining clean facilities. "The HCV is so fragile that it can be disinfected with soap and water."

> ***"This is a socioeconomic disease, and while 60% of infections happen in hospitals and clinics, the rest are down to bad social practices."***

The MOHP is taking action on this front. In 2008, it began a new campaign, with two major elements. "The first is education, where we regularly teach hospital staff about infection control," says Mohamed Ezz El-Arab, head of the tumour unit at NHTMI. "Second, every hospital now must have an infection control committee which has clear authority to monitor what is going on in the hospital."

Furthermore, HCV and infection control is now an important course that is taught in all medical universities, unlike the brief mention it received a few decades ago. There are also meant to be regular inspections of public hospitals and clinics to ensure that they meet the infection control standards set by the WHO and that proper waste-disposal protocols are in place.

Critics like Miller say that actions speak louder than words. He remains unconvinced of the programme's value or of its implementation. There have been recorded instances, he says, where public hospitals have dumped their waste "directly, without treatment, into the solid waste stream", violating WHO specifications. "This is a scandal."

Just last October, the Egyptian weekly magazine *Rose al-Yousef* reported that dialysis filters had been found in the non-medical waste disposal in front of a university hospital. Rubbish scavengers had been washing and reselling them for a fraction of their original price. In 2008, the Egyptian daily *Al-Masry Al-Youm* reported that truckloads of medical waste, including blood bags, had been delivered to garbage collectors for reuse.

Disposing of medical waste in this way is highly unsafe and risks passing on the virus. Medical waste disposal is "better controlled in Cairo now," says Ezz El-Arab. "But while we can follow-up on what is happening in Cairo, the problem is in rural areas further away where monitoring is hard."

THE REVOLUTION GOES VIRAL

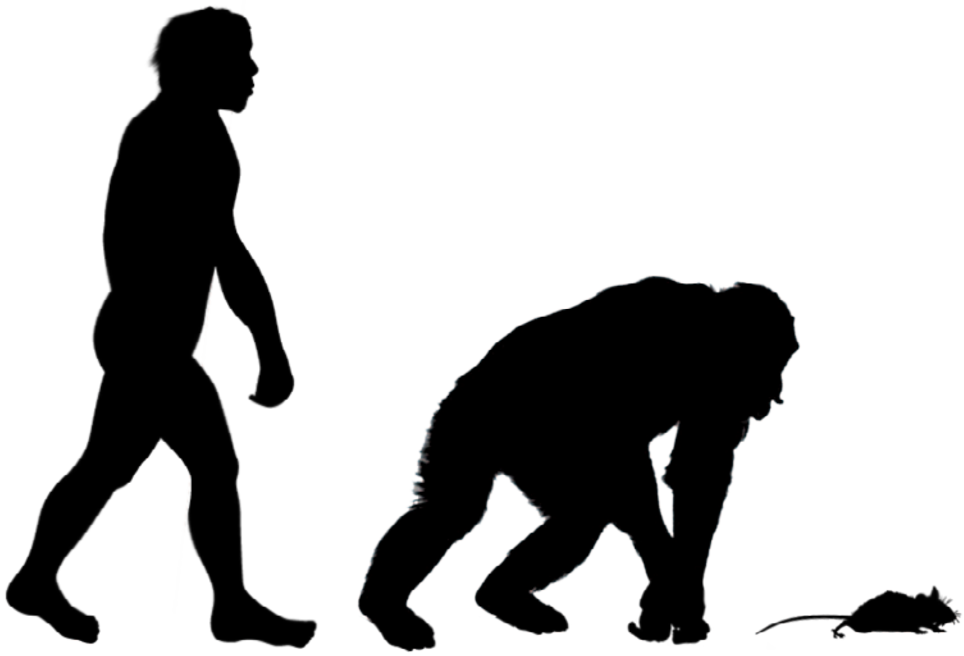
Measuring the success of these HCV programmes is hard because there are no objective data available. "Only those in the MOHP have any information on the status of this programme," Miller says. Without good data, it is not possible to verify whether the rate of new HCV infection in Egypt is decreasing. And even if there has been a drop in the rate, the overall level of infection is still much higher than anywhere else in the world. It can take 20 to 30 years of HCV infection before people develop serious complications. So the worst effects of the epidemic will surface over the next few years.

There is a ray of hope. The MOHP is part of the government established by the former President Hosni Mubarak. Miller suggests that, after the popular revolution in early 2011 that toppled Mubarak, there may be an opportunity for implementing better prevention programmes as many of the ministries undergo changes. "Post revolution means that HCV prevention should and can be accomplished by non-governmental, non-profit organizations," he says. "The MOHP has to come clean with its failure to prevent or reduce HCV transmission." At the same time as creating a new government, perhaps the revolution will also offer a uniquely Egyptian solution to the HCV epidemic. ■

Mohammed Yahia *is editor of* Nature Middle East.

1. El-Zanaty F. & Way, A. Egypt Demographic and Health Survey 2008 (Egyptian Ministry of Health, El-Zanaty and Associates and Macro International, 2009).
2. Miller, F. D. & Abu-Raddad, L. J. *Proc. Natl Acad. Sci. USA* **107**, 14757–14762 (2010).
3. Kamal, S. M. & Nasser, I. A. *Hepatology* **47**, 1371–1383 (2008).
4. Kamel, M. A. *et al. Lancet* **340**, 427 (1992).

PICTURE CREDIT



RESEARCH TECHNIQUE

The murine candidate

Small animals that mimic human hepatitis C infection will help researchers pinpoint weakness in the viral life cycle.

BY ELIE DOLGIN

The hepatitis C virus (HCV) is hard to study. Most of what researchers know about how it multiplies comes from cell-culture systems. Such cellular set-ups have proven invaluable for developing new drugs, including protease inhibitors and polymerase inhibitors, which prevent the virus from replicating its components inside the cell. Yet these cell-based systems fail to capture other important parts of the viral life cycle, such as the step before replication, when the virus attaches to liver cells and gains entry. What's more, cell-culture systems cannot reproduce the interaction between the immune system and the virus nor can they recapitulate entire organs so that liver pathology can be studied. For these reasons, researchers interested in how the virus causes disease have long sought a small-animal model.

Common laboratory animals, including rodents and most primates, are not susceptible hosts for HCV. Scientists have therefore had to settle for chimpanzees, which, like humans, are vulnerable to chronic HCV infections. However, "for ethical and economic reasons, the chimp is a terrible model," says Matthew Evans, an HCV researcher at the Mount Sinai School of Medicine in New York. Research involving chimpanzees is banned in many parts of the world, including Europe. And in most places where experimentation with great apes is allowed, laws against euthanizing chimps require investigators to fund the animals' long-term care — a prohibitively expensive commitment.

That's where a colony of ordinary-looking black mice running around in cages on the fourth floor of the Rockefeller University Comparative Bioscience Center in New York comes in. These animals might not look special, but they have been engineered to express either a pair or a quartet of human genes and, as such, are the first small animals with fully functioning immune systems that are prone to HCV infection. Using these models, "you can actually now look at hepatitis C virus entry *in vivo*," says Rockefeller immunologist Alexander Ploss, who developed the animals together with Charles Rice, executive and scientific director of the Center for the Study of Hepatitis C in New York.

➲ NATURE.COM
to read the latest research using animal models
go.nature.com/qDGI2e

These mice, and others like them, could provide a cheaper, more robust and less ethically fraught route to HCV drug and vaccine discovery.

THE HUMAN SIDE

Getting to this point has been a hard slog. In the years immediately after the virus was first described in 1989, many research teams developed transgenic mice carrying one or more genes encoding HCV proteins. Thus it was possible to study HCV-induced liver pathology without infecting mice with the virus. This approach still has some proponents. Last year, a team led by Matti Sällberg, a viral immunologist at the Karolinska Institute in Stockholm, used mice expressing the viral protease and showed that treatment with a drug targeting the cytokine tumour-necrosis factor-α led to improved liver function[1].

But the approach is highly artificial, leading to overexpression of the introduced viral genes and ignoring the rest of the viral life cycle. Over the past decade, most researchers have moved away from this set-up in favour of systems that involve infecting animals with the virus.

The first such model was reported ten years ago by a team led by transplant surgeon Norman Kneteman at the University of Alberta, in Edmonton, Canada. Kneteman's group engineered mice to express a gene that kills off the animals' own liver cells, which aren't susceptible to HCV infection; in their place they transplanted human liver cells, which are. These mice with humanized livers could be infected with HCV[2]. "This was the first [mouse] model that actually allowed HCV infection for prolonged periods of time by the normal route," says Kneteman.

These animals have proven useful for testing many candidate drugs. For example, a Japanese team led by Hiroshima University's Kazuaki Chayama treated Kneteman's liver transplant mice with a combination of new drugs: the protease inhibitor telaprevir (from Vertex Pharmaceuticals, based in Cambridge, Massachusetts) and the experimental polymerase inhibitor MK-0608 (from drug giant Merck, headquartered in Whitehouse Station, New Jersey). Late last year, the team reported that this combination eliminated the virus from the animals after a month of therapy and prevented the emergence of drug resistance, which often arises in mice and humans treated with either drug alone[3].

But to facilitate the human tissue transplant, the mice must be engineered to lack components of their immune system. The animals are thereby rendered poor models for testing drugs that alter the immune system, known as immunotherapies. Generating these mice also presents special difficulties. For one,

researchers can't breed chimaeric animals. And the mice are sickly because of the liver toxic gene.

Two recent transplant models of HCV infection provide improvements over Kneteman's mice[4,5]. Both types of mouse are less frail because of technical workarounds that allow researchers to introduce the liver deficit later in life. The model developed by Lishan Su, an immunologist at the University of North Carolina at Chapel Hill, in collaboration with Ploss and Rice at Rockefeller, also involves transplanting human blood stem cells into the animals to reconstitute a human-like immune system[5]. Of all of the published reports, says Su, "this is the only one that has both the immune system and the human liver in a chimaeric animal", creating a living platform for testing vaccines and immunotherapies in a human-like model.

Even though Su's mice generate a human T-cell response against the virus when infected, they still lack a complete immune system. "What we need now is a mouse — an immunocompetent, normal, mouse — that can be infected by a hepatitis C virus capable of replicating, spreading and initiating an immune response," says Frank Chisari, a virologist at The Scripps Research Institute in La Jolla, California. "We are light years away from that because that virus does not like to infect or replicate in mouse cells." But scientists are getting closer.

ENTRY LEVEL POSITION

To gain entry into liver cells, HCV hijacks four proteins. Although mice naturally produce these proteins, the human versions of two of them are needed for viral entry[6]. The black rodents at Rockefeller are the first animals into which the required human entry factors have successfully been introduced. "This has a lot of applications," says Ploss. "Right now, it's useful to measure HCV entry and potential entry inhibitors."

"This is a big advance," says Michael Houghton, a virologist at the University of Alberta, who co-discovered HCV more than 20 years ago. "It's been difficult to do vaccine research for hepatitis C because of the lack of an animal model other than the chimp. Now we can start using different vaccine strategies in mice to see which are best at eliciting a protective response."

Ploss's mice are the first such animals with a fully intact immune system that are susceptible to the viral infection. But the infection stops after cell entry: the virus does not seem to replicate. "You can recapitulate HCV entry," says Ploss, "but replication is still very inefficient and not detectable by conventional methods." So the big challenge now remains identifying whether additional human factors are needed to achieve the next step of the HCV life cycle in mice.

After replication comes assembly, when the viral components are gathered into new infectious particles that will be released from the cell and invade other cells. Fortunately, this final stage in the viral life cycle seems to be possible in mouse cells without introducing any human proteins, according to research presented at this year's International Liver Congress, in Berlin, by Ralf Bartenschlager, a molecular virologist at Heidelberg University in Germany. If the barriers to replication can be overcome, Bartenschlager says, it should be straightforward to get a full infection cycle going in a mouse. "We have the early steps; we have the late steps; the big black box now is the step in between."

It took more than a decade for scientists to deduce the factors needed for HCV cell entry. But Thomas Baumert, a hepatologist and virologist at the University of Strasbourg in France, is confident that the community will solve the problem of replication much faster. "We have better model systems now, so I think we can advance more rapidly." Within five years, he predicts, "it will be possible to produce transgenic mice for the entire viral life cycle".

Rice is equally confident this approach will work — but he is hedging his bets. Even such a model would have its drawbacks, he says, because the more mouse-like the model, the further removed it is from the human system. That's why even as his lab is aggressively pursuing a transgenic animal, he maintains active collaborations to develop other models, including new transplant chimaeric mice with humanized livers and immune systems. Other researchers are looking to animals that provide the natural susceptibility of primates without the ethical baggage (see 'The turn of the shrew'). "All of these things should be pursued in parallel," Rice says, "because we really don't know which of these models is going to be the best for a given application."

And so Rice and others continue to try and build a better mouse to help the research community beat a path to new HCV treatments. ■

THE TURN OF THE SHREW

Unusual model isn't persuading researchers of its practicality

Although most of the work developing small-animal models of hepatitis C virus (HCV) infection has focused on mice, some research teams have advanced an alternative model: the northern treeshrew (Tupaia belangeri). This squirrel-shaped animal shares a common ancestor with apes and is the only non-ape species known to be naturally susceptible to HCV. Last year, the first longitudinal analysis of HCV-infected tree shrews showed that, over the course of three years, the animals developed chronic hepatitis, fatty liver degeneration and liver cirrhosis[7]. "It's very similar to HCV infection in human beings," says study co-author Kyoko Tsukiyama-Kohara of Kumamoto University in Japan.

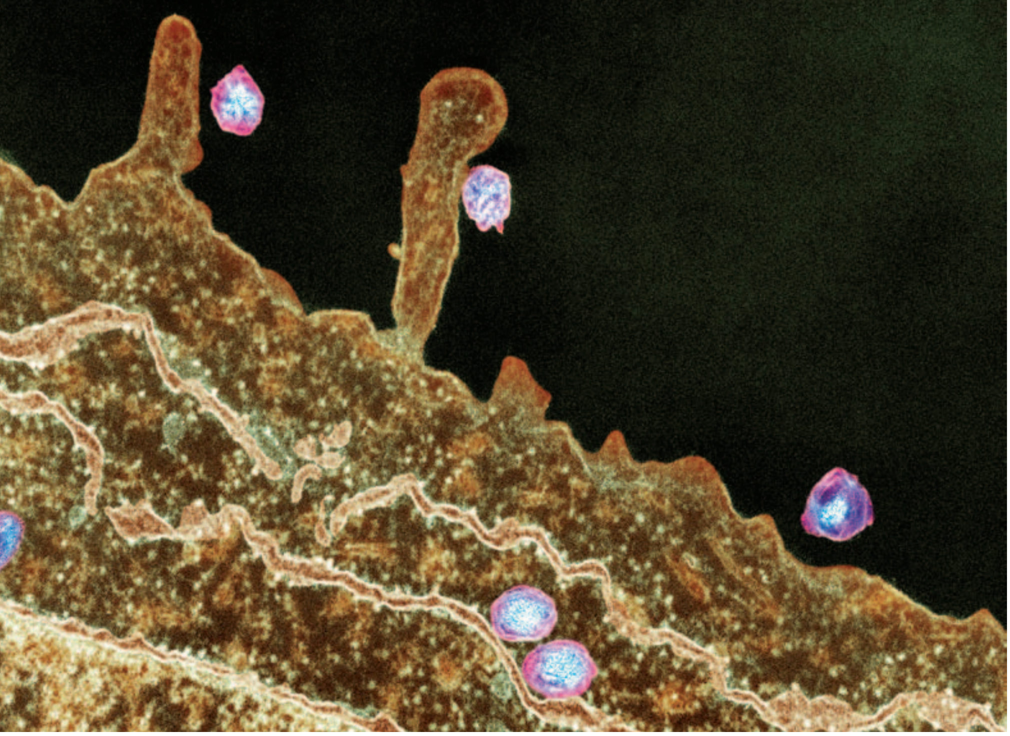
But few research teams have managed to establish long-term infections in the animals. And given the limited track record of tree shrews in drug discovery, most scientists agree that more traditional lab animal models of infection, such as mice, are needed. "If you're going to take a multimillion dollar drug and do your final trial before you go into humans, you need to have a reproducible model," says Robert Lanford, who has studied HCV in chimps for more than 20 years at the Texas Biomedical Research Institute in San Antonio.

NEIL BOWMAN/FLPA

The northern treeshrew, a natural host to hepatitis C virus, is proving an unpopular model of infection.

Elie Dolgin *is a news editor with* Nature Medicine *in New York.*

1. Brenndörfer, E. D. *et al. Hepatology* **52**, 1553–1563 (2010).
2. Mercer, D. F. *et al. Nature Med.* **7**, 927–933 (2001).
3. Ohara, E. *et al. J. Hepatol.* **54,** 872–878 (2011).
4. Bissig, K. D. *et al. J. Clin. Invest.* **120**, 924–930 (2010).
5. Washburn, M. L. *et al. Gastroenterology* **140**, 1334–1344 (2011).
6. Ploss, A. *et al. Nature* **457**, 882–886 (2009).
7. Amako, Y. *et al. J. Virol.* **84**, 303–311 (2010).



The hepatitis C virus attacks liver cells, which that can lead to cirrhosis and cancer.

THOMAS DEERINCK, NCMIR/SCIENCE/PHOTO LIBRARY.

VACCINES

A moving target

The hepatitis C virus has a set of cunning ways to evade immunity, but researchers are turning the immune system on it.

BY MICHAEL EISENSTEIN

The hepatitis C virus (HCV) dodges the immune system with devilish resourcefulness. Its vaunted genetic variability is an obvious advantage in escaping recognition by antibodies and T cells. But the virus also takes the fight directly to the host by actively suppressing the innate immune system, a key early mechanism in response to infection. In parallel, HCV uses diverse stratagems to mask its surface from antibodies that might neutralize infection. Recent evidence suggests that when all else fails, HCV hides — creeping directly from one liver cell to another, without exposing itself to the immune system[1]. "It's a special challenge as a vaccine target, and 75% of patients develop infection for life," says Michael Houghton, a virologist at the University of Alberta in Edmonton, Canada, who co-discovered HCV in 1989.

Given these facts, one might expect HCV vaccine research to be an exercise in futility. But the minority of patients who banish the virus during acute infection offer hope that the immune system might be coaxed into rising to the challenge. "If somebody has recovered from a primary HCV infection, any subsequent infection is much milder, and they tend to clear the virus far more quickly," says Marian Major, a virologist with the US Food and Drug Administration. "This protects the patient from chronic infection, which is the goal we need to meet for any HCV vaccine."

Accordingly, the most aggressively pursued — and clinically advanced — research has focused on the development of therapeutic vaccines that can beef up an otherwise inadequate immune response and treat chronic disease. Prophylactic vaccines are also in development, but they face economic, ethical and social problems, on top of daunting scientific challenges. However, both approaches need to overcome a common problem: HCV is a wily operator and will not be easy to beat.

CALLING IN FOR BACKUP

Several companies are actively testing therapeutic vaccine candidates, most of which use a non-pathogenic viral vector that presents HCV proteins to the immune system. For example, French biopharmaceutical company Transgene, based in Illkirch Graffenstaden, is adapting the vaccinia virus, used in the smallpox vaccine, to provoke an immune response against a trio of HCV targets. In early clinical trials, Transgene's TG4040 vaccine by itself achieved a modest reduction of viral load in chronically infected patients. Transgene recently started a phase II study to further assess TG4040's efficacy in combination with the standard antiviral regimen of interferon-α plus ribavirin.

➲ NATURE.COM
to read the latest news on hep C viral resistance
go.nature.com/xWxx06

Instead of using a vector, the Swedish company ChronTech Pharma, based in Huddinge, Sweden, vaccinates patients with naked DNA molecules encoding HCV genes. DNA vaccines are considerably cheaper to manufacture than vector-based vaccines, although their track record is patchy. But, by using an electrical current to pulse DNA directly into muscle tissue, ChronTech has been able to markedly improve DNA uptake while eliciting a strong local immune response. "For patients who went through the vaccine trial and were then put onto standard care, their virus disappeared very rapidly, and they also had an unusually high cure rate," says ChronTech co-founder Matti Sällberg, a viral immunologist at the Karolinska Institute in Stockholm. Indeed, five of six patients treated with ChronVac-C in conjunction with a standard drug regimen successfully eliminated HCV infection (relative to about half of all patients with existing drugs alone), and in March the company received approval for a phase II trial.

Both Transgene's and ChronTech's vaccines stimulate the cellular immune response — the first stage of the adaptive immune response, which kicks in after innate immunity (see 'HCV versus the immune system'). In the cellular immune response, killer T cells destroy infected liver cells when they recognize surface features, known as epitopes, that indicate the presence of virus. Cellular immunity seems to be the front line in preventing progression of acute infection. "Spontaneous resolvers are those who elicit early and broad cellular immune responses against multiple epitopes," says Houghton.

After the cellular immune response comes the humoral response, in which B cells secrete antibodies that can bind to and possibly neutralize HCV. The importance of humoral immunity in beating back acute HCV infection remains murky. "We need to learn a lot more about whether neutralizing antibodies are important," says Genevieve Inchauspé, head of the Infectious Diseases Department at Transgene. Helper and killer T-cell-oriented therapies have shown some effectiveness in humans, and there is evidence that HCV might encounter problems in eluding T-cell recognition[2]. Although the virus can acquire mutations that allow it to escape many immune-system traps, the mutations it requires to avoid T cells are detrimental and "actually reduce viral fitness", says Sällberg.

Yet even the best therapeutic vaccine candidates induce reactions that fall far short of those of spontaneous resolvers. Most ongoing trials therefore add these experimental therapies to the current standard treatment, but the existing regimen is notable for its nasty side effects. "The ultimate goal would be to be able to replace interferon and ribavirin with a vaccine," says Inchauspé, "or perhaps reduce the

therapy: make it shorter with fewer injections." She adds that patient interest is high enough that Transgene completed recruitment for its phase II trial ahead of schedule.

PREPARED FOR THE WORST

As safer and more effective drugs approach the marketplace, thus covering the therapeutic front in the battle against HCV, many vaccine developers are trying to address the traditional objective of vaccines: preventing disease from taking hold in the first place.

While working at the biotechnology firm Chiron, later acquired by Swiss pharmaceutical firm Novartis, Houghton managed the team that developed the first preventive HCV vaccine candidate, based on an injection of purified recombinant viral envelope proteins. This Chiron/Novartis vaccine was designed to elicit the production of antibodies while stimulating helper T cells, which in turn stimulate killer T cells and thus promote cellular immunity. In preclinical studies, significantly fewer vaccinated animals exhibited chronic infection relative to unvaccinated controls after exposure to HCV[3]. In phase I trials, human volunteers responded to vaccination with a strong helper T-cell response while producing detectable levels of antibodies with the capacity to neutralize diverse viral subtypes[4].

Although Novartis developed a plan for a phase II trial to assess the vaccine's protective efficacy, work lost momentum after Houghton's departure in 2007. Development was ultimately suspended, largely owing to financial considerations. "Although I believe it could work with 70–80% efficacy, it's a difficult and expensive vaccine to make," says Houghton. He would like to continue to develop and refine the approach in his new lab, with an eye towards future clinical trials.

The only other HCV vaccine candidates in the pipeline are from Okairos, a Rome-based biopharmaceutical company that was spun off from drug giant Merck, that recently embarked on phase I trials with two vaccine formulations. Okairos's prophylactic strategy uses two injections — an initial 'prime' and a follow-up 'boost' — that contain distinct viral vectors, each of which expresses the same set of HCV-derived proteins. This combination approach prevents the immune response from focusing its attention on the vector at the expense of the intended target, explains Alfredo Nicosia, chief scientific officer at Okairos. Okairos presented initial data from one trial at this year's International Liver Congress, in Berlin, showing that seven of ten vaccinated volunteers generated a broad T-cell response against peptides derived from a variety of HCV subtypes. This reaction was sustained over the course of a year following the prime injection, and Nicosia is optimistic that this vaccine will be ready for phase II trials later this year.

The motive for developing a prophylactic HCV vaccine is clear: to avoid the need for a toxic, unpleasant and expensive treatment that doesn't always work. A full course of ribavirin and interferon-α costs US$25,000 in the United States. Nevertheless, there is little evidence to suggest an imminent scramble for the largely empty prophylactic marketplace.

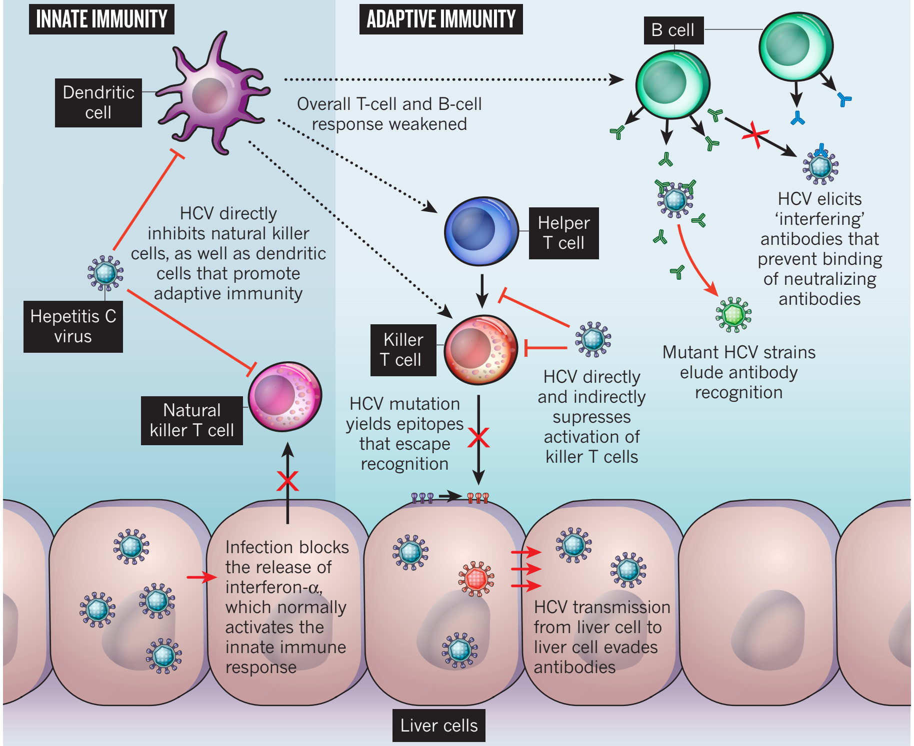
One reason may be that the advent of potent new antiviral drugs such as telaprevir (from Vertex Pharmaceuticals, based in Cambridge, Massachusetts) and boceprevir (from Merck, based in Whitehouse Station, New Jersey) offers an efficacy of HCV treatment that removes some of the urgency from developing preventive strategies. Furthermore, because of the relatively inefficient sexual transmission of the virus, most new cases of HCV arise from unsanitary medical practices in the developing world or among injection drug users; the limited financial resources of this target population could be a serious deterrent to companies that might otherwise consider embarking on the expensive and risky journey of vaccine development. Under these conditions, commercial success might depend on the willingness of health regulatory agencies to integrate HCV prevention into the standard vaccination arsenal. Houghton sees it as an open question whether those agencies would recommend a universal HCV vaccine for adolescents, "knowing that it's mainly the product of high-risk behaviours like intravenous drug use."

Even if a vaccine manages to clear those hurdles, navigating through the testing process will be more tricky than usual. HCV is so good at evading the innate immune system that a successful prophylactic vaccine might not prevent acute infection, although, by bolstering adaptive immunity, the vaccine should prevent a chronic infection and subsequent reinfection. This risk of acute infection will add an ethical complication to clinical trials involving prophylactic treatment of patients in high-risk groups. Doctors who monitor these trials would be obliged to warn people who develop an acute infection and offer them prompt access to treatment, directly confounding the study results.

These challenges are not insurmountable and, as a positive example, Houghton points to the vaccine against hepatitis B virus, which also disproportionately affects injection-drug users. He adds that the expertise of investigators at agencies such as the National Institute on Drug Abuse, in Bethesda, Maryland, has been invaluable in HCV vaccine trial planning and recruitment. However, it is important to manage expectations: even a successful vaccine will probably achieve only an efficacy of 70–80% because of HCV's remarkable escape artistry. "It's a race between the vaccine-induced immune response and the virus," says Nicosia, "and we need to tilt it in favour of the former in order to achieve prevention." ■

HCV VERSUS THE IMMUNE SYSTEM

The many routes (red) by which the hepatitis C virus suppresses and evades the human immune response (black).



Michael Eisenstein *is a journalist based in Philadelphia, Pennsylvania.*

1. Brimacombe, C. L. *et al. J. Virol.* **85**, 596–605 (2011).
2. Söderholm, J. *et al. Gut* **55**, 266–274 (2006).
3. Houghton, M. *Immunol. Rev.* **239**, 99–108 (2011).
4. Frey, S. E. *et al. Vaccine* **28**, 6367–6373 (2010).

Protesters outside a US Department of Health and Human Services building in Alameda County, California.

OASIS CLINIC

PERSPECTIVE

Test and treat this silent killer

The scourge of hepatitis C virus in the United States is woefully underestimated. **Brian R. Edlin** reckons it's time the infection is given the priority it demands.

At least five million people in the United States have been infected with the hepatitis C virus (HCV)[1]. That's about five times as many people as are infected with HIV. A sound public-health response to the HCV epidemic requires urgent investment in prevention, testing, treatment and research — investments that were made for HIV — to avert greater financial costs and loss of life. Indeed, a January 2010 report by the Institute of Medicine (IOM), part of the US National Academy of Sciences, warned that viral hepatitis will remain out of control unless adequate resources are devoted to prevention, control and surveillance[2].

This is not the response that we are witnessing. The US government has all but ignored the threat of HCV and is underfunding prevention, treatment and research into the disease (see 'US Response to the HIV and viral hepatitis epidemics'). The Action Plan to Prevent, Care and Treat Viral Hepatitis, which was released in May 2011 by the US Department of Health and Human Services in response to the IOM report, does not include an intention to increase funding for viral hepatitis.

Control of any epidemic starts with an accurate understanding of the magnitude of the problem, but the scope of the HCV epidemic in the United States is poorly understood. For example, the Centers for Disease Control and Prevention (CDC) estimate of the prevalence of HCV infection in the United States is four million people[3]. But this relies on data from a national household survey that has long been known to suffer from non-response bias and to exclude high-risk populations such as homeless people and prisoners. This survey underestimated[4] the prevalence of HIV infection in the United States by a factor of 1.4 to 2.0. If the HCV estimate is similarly biased, then 6–8 million Americans are likely to have been infected with HCV.

Even less is known about the rate at which HCV is spreading. The CDC assumes that there are about 20 new infections for each case reported to its surveillance system and estimates that about 18,000 new infections occur each year — about one every 30 minutes. The actual number could be much higher. Investigators in Seattle and Baltimore, and my research teams in San Francisco and New York, have found that injection drug users, in whom most new infections in the United States occur, rarely experience symptoms when they acquire HCV — and when they do, they rarely seek medical attention[5]. Probably fewer than 1 in 100 new infections in injection drug users, and possibly considerably fewer, are reported to health departments. Infections in this entire group of people are thus almost invisible. What's more, the recent explosion of opioid use among young adults has created a new silent HCV epidemic. In Massachusetts alone, more than 1,000 new HCV infections among 15–25 year olds annually since 2007 have been linked to this recent wave of drug use[6].

The CDC's surveillance system covers people who are not homeless or institutionalized, have nothing to fear from the authorities, have access to health care, feel ill when they contract HCV infection and visit a doctor when they feel ill. But HCV disproportionately affects groups for whom those attributes often don't hold true (see 'Hepatitis C is a disease of the marginalized').

Even if the extent of the epidemic is unclear, basic prevention strategies could still be implemented. Prevention is the cornerstone of disease control, so the lack of funding in this area is especially troubling. Some have argued that HCV infection cannot be prevented. But there is strong evidence that the interventions that decreased HIV transmission among drug users — community-based outreach and education, testing and counselling, access to sterile syringes and substance-use treatment — also markedly reduced HCV transmission[7]. Once drug users learned about how HIV was spread, and barriers to obtaining sterile injection equipment were lowered, users eagerly adopted safer injection techniques, and both HIV and HCV incidence rates fell by as much as an order of magnitude from the late 1980s to the late 1990s.

Liver disease rates will continue to rise, and about 150,000 Americans will die from viral hepatitis in the next decade.

HCV transmission rates were, however, far higher than those for HIV to begin with, and they remain unacceptably high. But HCV incidence rates in injection drug users fell from 15–20% per month in the 1980s to 10–25% per year after HIV prevention interventions were introduced. The median time between the first use of injection drugs and HCV infection rose from 3–4 months in the 1980s to 4–7 years in the

SOURCE: REFS 1, 2, 3, 7. AND BRIAN R. EDLIN, UNPUBLISHED DATA.

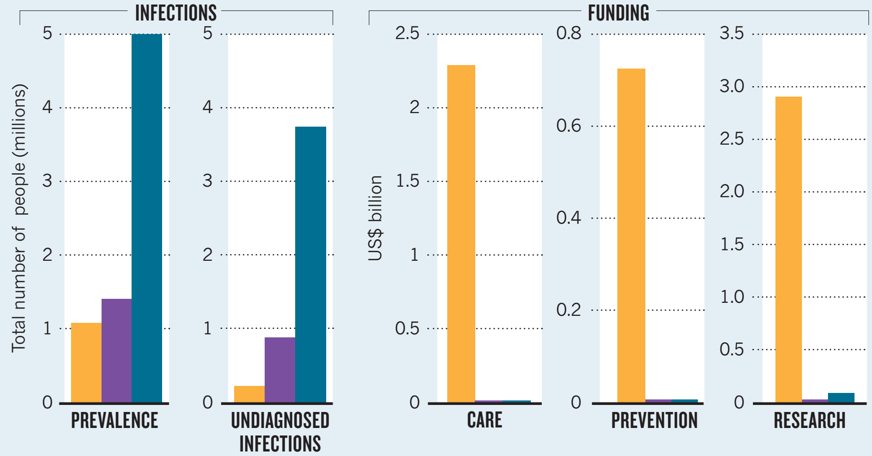
PLAN OF ACTION

Confronting the epidemic

PREVENTION: Provide needle-exchange facilities, syringe access, community-based outreach and education, community-based services, testing and counselling, links to care and substance-use treatment to all those in need.
TESTING: Provide hepatitis C testing to populations with an elevated prevalence wherever HIV testing is available.
CARE: Provide multidisciplinary services, comprehensive and continuing primary and specialty medical care, substance-abuse treatment, mental healthcare, case management, support services, and provider education and training nationally to all those in need.
PRISONS: Provide HCV prevention, testing and treatment services in correctional institutions.
SURVEILLANCE: Develop and implement surveillance systems for those currently overlooked by current systems.
RESEARCH: Prioritize the development of better interventions so that prevention, testing and care can be more effectively and efficiently provided to those who need it.

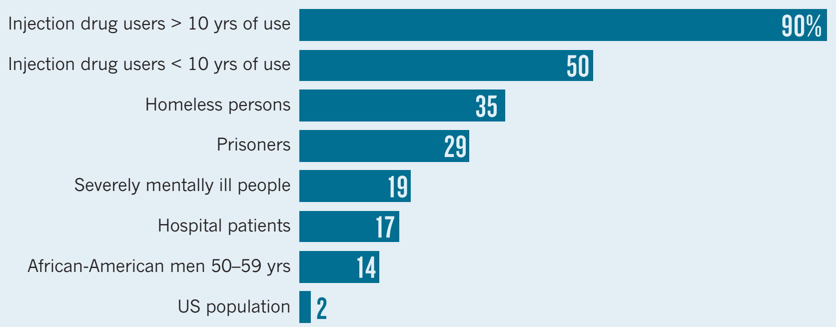
US RESPONSE TO HIV AND VIRAL HEPATITIS EPIDEMICS

Hepatitis C infection is at least five times more prevalent as HIV infection in the United States, yet funding lags far behind.



HEPATITIS C IS A DISEASE OF THE MARGINALIZED

Hepatitis C disproportionately affects groups who are under-represented in health surveillance systems and who are underserved by the healthcare system. Percentage of each group testing positive for HCV infection.



1990s, opening a window of opportunity in which to slow the spread of the virus. But effective preventive interventions are not available in many locations and are poorly funded where they do exist. And research on new interventions to reduce transmission further, which is urgently needed, is not a priority at the National Institutes of Health (NIH).

About half of all HCV infections can be cured with a single 6–12-month course of therapy, and new drug regimens will be more effective. But the resources to provide testing and deliver care are missing. US government policy recommends that all adults be screened for HIV, and there has been significant progress towards that goal: 83 million people have been tested, and CDC grants provide funding for another 1.4 million tests each year. But no such policy or funding supports HCV testing — even though more than ten times as many Americans have undiagnosed HCV as have undiagnosed HIV infection, and half or more of them could be cured with a single 6–12 month course of therapy. Moreover, public-health programmes stop at the doors of most correctional facilities, abandoning the infected and those at risk at the moment when providing prevention and treatment services would be most practical. In 2009, the IOM put the development of better treatment strategies for at-risk populations such as injection drug users on its list of the nation's top research priorities[8] — but the NIH has yet to respond to the call. As a result, liver disease rates will continue to rise, and about 150,000 Americans will die from viral hepatitis in the next decade[2]. Public and private healthcare expenditure on hepatitis C, which is estimated at US$30 billion today, is expected to exceed US$80 billion, with an increasing proportion of the cost falling to public sources[9]. If we fail to act now, these expenses will overwhelm our already overburdened public healthcare system.

The United States launched a vibrant response to the HIV/AIDS epidemic and made great advances in surveillance, prevention, care and treatment by investing in research and services in these areas. A timely, proportionate response to HCV will require leadership and resources. Nothing less will mitigate the extraordinary toll that viral hepatitis could take on the public's health and on the healthcare system. ■

➲ NATURE.COM
For supplementary information on this article
go.nature.com/7xF13KO

Brian R. Edlin *is a professor of medicine at the State University of New York Downstate College of Medicine in New York.*
e-mail: brian.edlin@downstate.edu

1. Edlin, B. R. *Hepatology* **42**, 213A (2005).
2. Institute of Medicine. Hepatitis and Liver Cancer: A National Strategy for Prevention and Control of Hepatitis B and C (National Academies Press, 2010).
3. Armstrong, G. L. *et al. Ann. Intern. Med.* **144**, 705–714 (2006).
4. McQuillan, G. M., Khare, M., Karon, J. M., Schable, C. A. & Vlahov, D. J. *Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **14**, 355–360 (1997).
5. Edlin, B. R. & Carden, M. R. *Clin. Infect. Dis.* **42**, 673–676 (2006).
6. Centers for Disease Control and Prevention. *MMWR Morb. Mortal. Wkly Rep.* **60**, 537–541 (2011).
7. Tseng, F. C. et al. *Hepatology* **46**, 666–671 (2007).
8. Institute of Medicine. Initial National Priorities for Comparative Effectiveness Research (National Academies Press, 2009), page 126.
9. Pyenson B, et al. Consequences of Hepatitis C Virus: Costs of a Baby Boomer Epidemic of Liver Disease. New York: Milliman (2009).

ILLUSTRATION BY SOPHIE CASSON/TREEINABOX.COM

DIAGNOSTICS

A testing journey

Results ready in minutes and more efficacious drugs will help find and treat the hundreds of millions of carriers.

BY SARAH DEWEERDT

Obtaining an accurate diagnosis of hepatitis C often means travelling a long road, with many stops along the way. First, you'll need a blood test to check for antibodies to the hepatitis C virus (HCV), an analysis that can take as long as two weeks. But even with antibodies, you might not be infected: 15–25% of exposed people clear the virus without needing any treatment.

The next step is to determine whether you have an active infection, and that means another blood draw to test for the presence of viral RNA — and another week or two of waiting for results. If that test comes back positive, you may need additional blood tests to determine your genotype (see 'Playing the odds') and the genotype of the virus, both of which affect how you are likely to respond to treatment.

This drawn-out journey is only part of the problem. HCV infection is often 'silent', with symptoms not developing for twenty years or more. Consequently, many infected people are unaware they have hepatitis C and never seek medical help. Estimates are that fewer than half of the people who are chronically infected know they have HCV. For some, this will not be a problem as the disease will not progress; for others it could be fatal.

Scientists and public-health advocates are now turning their attention to developing tests that are faster, cheaper and less invasive. They are also considering how to design screening programmes to identify people who might, unknowingly, be harbouring the virus. "It doesn't matter how good your treatments are if the majority of your affected population is not diagnosed," says Gregory Dore, head of the Viral Hepatitis Clinical Research Program at the University of New South Wales in Sydney, Australia. The good news is that hepatitis C can often be treated. And as improved medications with greater efficacy and shorter treatment times reach the market, there is a greater imperative to find and treat the people who are infected.

As many as 1 in 33 baby boomers might be infected.

QUICK AND CHEAP

Several companies have rapid antibody tests for HCV in the pipeline. The first one to hit the market was developed by OraSure Technologies, a biotechnology firm in Bethlehem, Pennsylvania. The company's OraQuick HCV Rapid Antibody Test was approved for use in Europe at the end of 2009 and in the United States in mid-2010, and is gradually being rolled out in these regions.

In the United States, the OraQuick test can be carried out with blood from a finger prick. In Europe, the less invasive method of taking a mouth swab for saliva has also been approved. Either way, a health professional places the sample into a vial and inserts a plastic strip that looks like a home pregnancy test: two red-purple lines means that the patient has antibodies to HCV[1]. The results are ready in 20 minutes.

"We believe it's going to dramatically expand testing opportunities," says Stephen Lee, OraSure's chief science officer. "Testing will be able to go on in a broad range of settings — doctors' offices, community outreach centres, needle-exchange clinics, even mobile testing vans." Similar rapid tests have greatly increased HIV screening and diagnosis over the last decade.

The next step in the diagnostic journey, testing for viral RNA, typically involves a method such as the polymerase chain reaction (PCR) to amplify the virus's genetic material to a detectable level. This step is time-consuming and expensive — much of the reason why antibody testing is usually carried out first. "One of the big issues is making RNA testing affordable," says John Ward, director of the Division of Viral Hepatitis at the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. Faster RNA testing, he says, would also be "a fantastic development".

One approach to improving RNA testing comes from a team led by Hassan Azzazy, a chemist at the American University in Cairo. "We are going directly to detecting HCV RNA in a single reaction, without amplifying the RNA of the virus," he says.

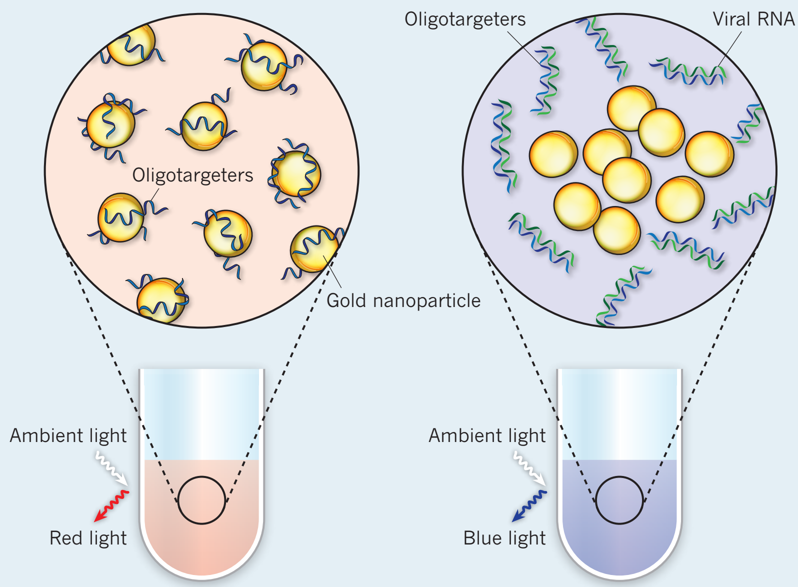
Azzazy's test uses tiny particles of gold, each about 15 nanometres in diameter[2]. Gold nanoparticles have an unusual optical property known as 'surface plasmon resonance': when the particles are distributed evenly throughout a liquid, they reflect light in a way that makes them appear red; however, when they clump together, they look blue.

To perform the test, Azzazy first takes serum from the patient and treats it to extract RNA from any virus present. To this treated serum he adds short pieces of DNA that are

complementary to HCV RNA. The gold nanoparticles are then added to this solution. In the absence of HCV RNA, the primers stick to the gold nanoparticles and separate them — and the solution appears red. If the virus is present, the primers pair with the viral RNA instead and the gold nanoparticles aggregate, turning the solution blue (see 'Red light, blue light'). The reaction can be carried out in a test tube, and the whole process takes about 30 minutes. "We don't need fancy infrastructure to run the test or to interpret the results," Azzazy says. It should cost about one-seventh as much as the current HCV RNA tests, he says, and could make antibody testing unnecessary.

RED LIGHT, BLUE LIGHT

Gold nanoparticles are added to a test tube containing a patient's serum and oligotargeters — short pieces of DNA designed to match genetic material from the hepatitis C virus (HCV). If there is no HCV, the oligotargeters stick to the nanoparticles and separate them and the solution appears red. If HCV is present, the oligotargeters pair with the viral RNA and the gold nanoparticles clump, turning the solution blue.



ONE-STOP TEST

Researchers at Brazil's Carlos Chagas Institute, part of the Oswaldo Cruz Foundation in Rio de Janeiro, aim to streamline the diagnostic process even further. Marco Krieger and colleagues say that a single RNA test could suffice both to diagnose HCV infection and to identify which subtype of the virus is present, which will help guide treatment.

There are at least six genotypes of HCV, most of which have several subtypes. In Brazil, subtypes 1a, 1b, 2a, 2b, 2c and 3a account for nearly all HCV infections. Because different genotypes are present in different parts of the world, a diagnostic test designed for one country might not work in another, says Marco Krieger, who is leading the project.

His group's test makes use of a technology called a liquid microarray, in which short pieces of DNA are attached to tiny plastic beads, or microspheres, floating in a sample[3]. It's similar to a gene chip in that it can detect many nucleic acid sequences simultaneously, which makes the test both rapid and flexible.

Currently, the test is designed to identify both genetic sequences that are common to most viral subtypes, and variable sequences that distinguish between the subtypes. Unlike current RNA tests, the method can detect many different sequences in a single reaction, and Krieger says that it could easily be adapted to detect a different set of genotypes. "The assay flexibility also allows us to add new sequences in the future, for instance, sequences related to drug resistance if necessary," he says.

SCREENING SELECTION

As more diagnostic technologies are developed, other issues arise. Ward points out the need for standards to ensure that results of different tests can be compared. The CDC has begun discussions with the World Health Organization about collaborating to create such guidelines.

Another equally important issue is how to apply the tests. In the United States and many other developed countries, where the overall prevalence of HCV infection is low, mass screening isn't cost effective. The CDC's current HCV screening recommendations, developed in the late 1990s, advise testing those known to have been at risk of exposure to the virus. But many people who contracted the virus several decades ago, perhaps through a blood transfusion or youthful experimentation with drugs, don't realize they are at risk and do not come forward to be tested.

According to researchers at the Amsterdam Public Health Service, the Internet could be used to find these people and help them. A team led by Maria Prins designed and validated an online screening questionnaire about risk factors such as former injection drug use, blood transfusions before blood bank screening began and immigration from countries with a high prevalence of HCV infection[4].

A mass media campaign including television and Internet advertisements brought almost 41,000 people to the website (www.heptest.nl), and 9,653 completed the questionnaire. Of the 1,480 people who were found to be at risk and eligible for a free blood test, 420 — nearly 30% — were tested. Prins was encouraged by this participation rate; in a similar effort aimed at syphilis, only 10% of those identified as at risk followed through and had a test, she says.

Study participants said they liked the convenience and anonymity offered by the online process. The biggest hurdle was encouraging people to take the questionnaire in the first place. "People don't see hepatitis C as something very threatening, and they confuse it with hepatitis B," says Prins. "They think they have already been vaccinated." In the future, the team may combine screening efforts for the hepatitis B virus and HCV.

To avoid a looming crisis, however, public health authorities will need to cast a wider net and not wait for people to seek out a test. In particular, it makes sense to target certain age groups. In developed countries, the peak of HCV transmission was between the 1960s and 1980s. Many of those infected during this period will soon begin to develop liver damage and possibly cancer: in the United States, as many as 1 in 33 baby boomers might be infected. According to a computer model designed by a team led by Lisa McGarry of i3 Innovus, part of the health-care information technology company Ingenix, in Medford, Massachusetts, screening everyone born between 1945 and 1970 would mean carrying out more than 80 million tests in the United States alone — but it could prevent 50,000 deaths. In fact, new CDC guidelines due to be released in 2012, might include a similar recommendation for age-based screening for HCV.

"A routine test in that age group is what you want," says Eugene Schiff, director of the Center for Liver Diseases at the University of Miami in Florida. He argues that this greatly expanded screening is key to making the journey to diagnosis shorter — and the burden of hepatitis C lower. "You're not going to get it done otherwise." ■

Sarah DeWeerdt *is a science writer in Seattle, Washington.*

1. Lee, S. R. *et al. J. Virol. Methods* **172**, 27–31 (2011).
2. Shawky, S. M., Bald, D. & Azzazy, H. M. *Clin. Biochem.* **43**, 1163–1168 (2010).
3. Duarte, C. A. *et al. PLoS ONE* **5**, e12822 (2010).
4. Zuure, F. *et al. Euro Surveill.* **15**, 19539 (2010).
5. Zuure, F. R. *et al. Am. J. Prev. Med.* **40**, 345–352 (2011).

The electromotive force of MnAs nanoparticles

ARISING FROM P. N. Hai, S. Ohya, M. Tanaka, S. E. Barnes & S. Maekawa. *Nature* **458,** 489–492 (2009)

Magnetic tunnel junctions can produce large magnetoresistance effects that are of use in a variety of applications. Hai *et al.*[1] recently published a very interesting paper in which the application of a large static magnetic field to a tunnel junction containing superparamagnetic MnAs nanoparticles resulted in the generation of an electromotive force. The authors attributed this phenomenon to a conversion of the nanoparticles' magnetic energy to electrical energy by way of quantum tunnelling. Here I point out that the electrical energy output measured by Hai *et al.*[1] was more than 1,000 times greater than the maximum amount of magnetic energy that could be induced in their MnAs nanoparticles by the applied magnetic field. Therefore the induced magnetic energy cannot be the source for the observed electromotive force, as was asserted by Hai *et al.*[1].

The maximum magnetic energy that can be induced in the MnAs nanoparticles by the applied magnetic field has the form $E_{\mathrm{mag}} = 2N\mu B$, where N is the number of nanoparticles in the sample, μ is the average saturation magnetic moment per nanoparticle, and B is the applied magnetic field. The values of these parameters as determined by Hai *et al.*[1] were $N \approx 10^9$, $B = 10\,\mathrm{kG}$ and $\mu = 2\mu_{\mathrm{B}}S$, where S is the average spin per nanoparticle (in units of $\hbar$), $\sim$200, and μ_{B} is the Bohr magneton. This yields $E_{\mathrm{mag}} \approx 7 \times 10^{-12}\,\mathrm{J}$ for their device. However, by using the inset of figure 2c in ref. 1 to calculate the electrical output power delivered to a $200\,\mathrm{k}\Omega$ load resistor and integrating over time, it can be seen that the electrical energy output of the device was greater than $10^{-8}\,\mathrm{J}$; this is more than a factor of 1,000 greater than E_{mag}.

D. C. Ralph[1,2]
[1]Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA.
[2]Kavli Institute at Cornell, Cornell University, Ithaca, New York 14853, USA.
e-mail: ralph@ccmr.cornell.edu

Received 27 January; accepted 5 April 2011.

1. Hai, P. N., Ohya, S., Tanaka, M., Barnes, S. E. & Maekawa, S. Electromotive force and huge magnetoresistance in magnetic tunnel junctions. *Nature* **458,** 489–492 (2009).

Competing Interests: declared none.

doi:10.1038/nature10142

doi:10.1038/nature10090

X-ray illumination of the ejecta of supernova 1987A

J. Larsson[1], C. Fransson[1], G. Östlin[1], P. Gröningsson[1], A. Jerkstrand[1], C. Kozma[1], J. Sollerman[1], P. Challis[2], R. P. Kirshner[2], R. A. Chevalier[3], K. Heng[4], R. McCray[5], N. B. Suntzeff[6], P. Bouchet[7], A. Crotts[8], J. Danziger[9], E. Dwek[10], K. France[11], P. M. Garnavich[12], S. S. Lawrence[13], B. Leibundgut[14], P. Lundqvist[1], N. Panagia[15,16,17], C. S. J. Pun[18], N. Smith[19], G. Sonneborn[10], L. Wang[20] & J. C. Wheeler[21]

When a massive star explodes as a supernova, substantial amounts of radioactive elements—primarily ^{56}Ni, ^{57}Ni and ^{44}Ti—are produced[1]. After the initial flash of light from shock heating, the fading light emitted by the supernova is due to the decay of these elements[2]. However, after decades, the energy powering a supernova remnant comes from the shock interaction between the ejecta and the surrounding medium[3]. The transition to this phase has hitherto not been observed: supernovae occur too infrequently in the Milky Way to provide a young example, and extragalactic supernovae are generally too faint and too small. Here we report observations that show this transition in the supernova SN 1987A in the Large Magellanic Cloud. From 1994 to 2001, the ejecta faded owing to radioactive decay of ^{44}Ti as predicted. Then the flux started to increase, more than doubling by the end of 2009. We show that this increase is the result of heat deposited by X-rays produced as the ejecta interacts with the surrounding material. In time, the X-rays will penetrate farther into the ejecta, enabling us to analyse the structure and chemistry of the vanished star.

Owing to the proximity of SN 1987A (located only 160,000 light yr away), we can study the evolution of the supernova in great detail. The central ejecta are surrounded by a ring of circumstellar material (Fig. 1) that was shed from the star 20,000 yr before the explosion in 1987[4]. Since the explosion, the ejecta have been expanding, and now the outer parts of the ejecta are colliding with the ring, causing it to brighten at all wavelengths[5–8]. The dense, central part of the ejecta contains most of the mass from the disrupted star and acts as a calorimeter for the energy input to the supernova. We have determined the energy input by tracking the energy output with the NASA Hubble Space Telescope (HST).

Because the ejecta are roughly elliptical in projection on the sky, we used an elliptical aperture to measure the brightness. To monitor a constant mass of expanding material, we allowed the measuring aperture to expand linearly with time. The axes of the aperture were therefore three times larger in 2009 than in 1994 (Fig. 1). Using this aperture, we determined the R- and B-band light curves of the ejecta, as shown in Fig. 2 (see Supplementary Table 1 and Supplementary Information, section 1, for further details of the observations and light curves). Our measurements show that the flux from the ejecta decays during the first ~5,000 d after the explosion, as expected from radioactive input, but then starts to increase, reaching a level that is two to three times higher at around day 8,000 (end of 2009). A new energy source must be present in addition to radioactive decay. Below, we consider a model for the declining phase and then discuss the new energy source that is responsible for the observed increase in flux.

The energy input to the declining phase of the light curve after ~1,500 d is expected to come from positrons produced in the decay of ^{44}Ti (refs 2, 9, 10). To test this, we use a model[2] with abundances taken from the 14E1 explosion model[11] and a ^{44}Ti mass of $1.4 \times 10^{-4} M_\odot$ (ref. 12; Supplementary Information, section 3), where $M_\odot$ is the solar mass. The model is shown in Fig. 3 together with the observed broad-band luminosities. The good agreement with the observations up to day 5,000 confirms that the ^{44}Ti positrons provide the energy input up to this point. However, after day 5,000 the model fails to describe the light curve; radioactive decay cannot explain the increase in flux that we observe.

One possible origin for the flux increase is the reverse shock that results from the interaction between the ejecta and the H II region inside the ring[13–16]. The reverse shock produces strong Lyα and Hα emission, which increased by a factor of ~1.7 between 2004 and 2010[16]. Although most of this emission originates close to the ring, there is also a component of projected high-velocity Hα emission that can be traced to the central parts of the ejecta[16] and which would therefore contribute to the flux we measure. To determine the contribution of the reverse shock to our light curves, we have examined HST Space Telescope Imaging Spectrograph spectra from 2004 and 2010 (Supplementary Information, section 2, and Supplementary Fig. 5). The reverse shock can be isolated in the spectra because of its boxy line profile, allowing us to place a limit on its contribution at $\lesssim$20%. Furthermore, this changes only marginally between 2004 and 2010, as the expanding measuring aperture remains well inside the area where most of the shock emission is seen. Importantly, an increase in flux is also seen in the [Ca II] doublet lines at rest wavelengths 7,292 Å and 7,324 Å between 2000 and 2010 (determined from Ultraviolet and Visual Echelle Spectrograph observations at the Very Large Telescope; Fig. 2). These lines have speeds of $\lesssim$5,000 km s^{-1}, implying that they originate in the inner ejecta (the projected ejecta speed near the edge of the ring is $\gtrsim$7,000 km s^{-1} at the present time). We conclude that the increase in flux occurs primarily in the inner ejecta and cannot be explained by emission from the shock region.

We believe that the strong X-ray flux produced in the ring collision is the dominant source of energy input to the ejecta. The X-ray flux from the ring increased by a factor of ~3 in the 0.5–10-keV band between day 6,000 and day 8,000[6], similar to what we find for the optical emission from the ejecta. To investigate this, we calculated the fraction of X-rays absorbed by the ejecta from a point source located at the ring, using the partially mixed 14E1 explosion model[17]. As shown in Supplementary Fig. 6, most of the observed X-ray flux is

[1]Department of Astronomy, The Oskar Klein Centre, Stockholm University, 106 91 Stockholm, Sweden. [2]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, MS-19, Cambridge, Massachusetts 02138, USA. [3]Department of Astronomy, University of Virginia, Charlottesville, Virginia 22903, USA. [4]Eidgenössische Technische Hochschule Zürich, Institute for Astronomy, Wolfgang-Pauli-Strasse 27, CH-8093 Zürich, Switzerland. [5]JILA, University of Colorado, Boulder, Colorado 803090-0440, USA. [6]George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, Department of Physics and Astronomy, College Station, Texas 77843, USA. [7]DSM/IRFU/Service d'Astrophysique Commissariat á l'Energie Atomique et aux Energies Alternatives, Saclay, Orme des Merisiers, FR 91191 Gif-sur-Yvette, France. [8]Department of Astronomy, Mail Code 5240, Columbia University, 550 West 120th Street, New York, New York 10027, USA. [9]Osservatorio Astronomico di Trieste, Via Tiepolo 11, Trieste 34131, Italy. [10]NASA Goddard Space Flight Center, Code 665, Greenbelt, Maryland 20771, USA. [11]Center for Astrophysics and Space Astronomy, University of Colorado, Boulder, Colorado 80309, USA. [12]225 Nieuwland Science, University of Notre Dame, Notre Dame, Indiana 46556-5670, USA. [13]Department of Physics and Astronomy, Hofstra University, Hempstead, New York 11549, USA. [14]ESO, Karl-Schwarzschild-Strasse 2, 85748 Garching, Germany. [15]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. [16]INAF-CT, Osservatorio Astrofisico di Catania, Via S. Sofia 78, I-95123 Catania, Italy. [17]Supernova Limited, OYV #131, Northsound Road, Virgin Gorda, British Virgin Islands. [18]Department of Physics, University of Hong Kong, Pok Fu Lam Road, Hong Kong, China. [19]Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA. [20]Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843-4242, USA. [21]Department of Astronomy, University of Texas, Austin, Texas 78712-0259, USA.

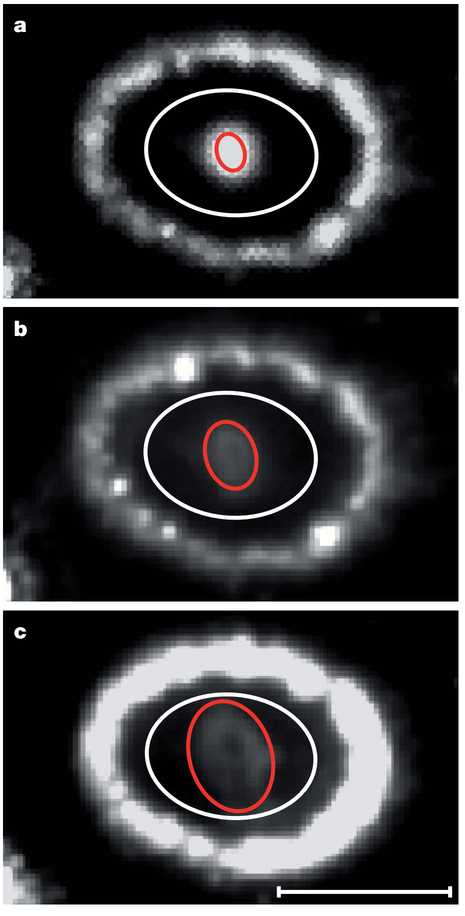


Figure 1 | HST R-band images. The observing dates are 1994 September 24 (**a**), 2000 November 13 (**b**) and 2009 April 29 (**c**), which correspond to 2,770, 5,012 and 8,101 d after the explosion, respectively. The scale bar in **c** represents 1′′. The circumstellar ring is inclined at an angle of 45° with respect to the line of sight and is approximately 1.3 light yr across. The red ellipse shows the expanding aperture used for the light curve in Fig. 2. By using an initial semi-major axis of 0.11′′ for the observation in 1994, we always follow the bright, central part of the ejecta without being significantly affected by emission from the circumstellar ring. The white ellipse shows the fixed aperture used for one of the light curves in Supplementary Fig. 2. The R-band emission from the ejecta is dominated by Hα emission with a small contribution from [Ca I] and [Ca II] lines, whereas the B band (Supplementary Fig. 1) is dominated by H I, Fe I and Fe II lines[12,22]. Only the densest, central parts of the ejecta are visible, owing to the low surface brightness of the outer parts. In reality, the ejecta extend to the ring, as is evident from the strong interaction with the ring.

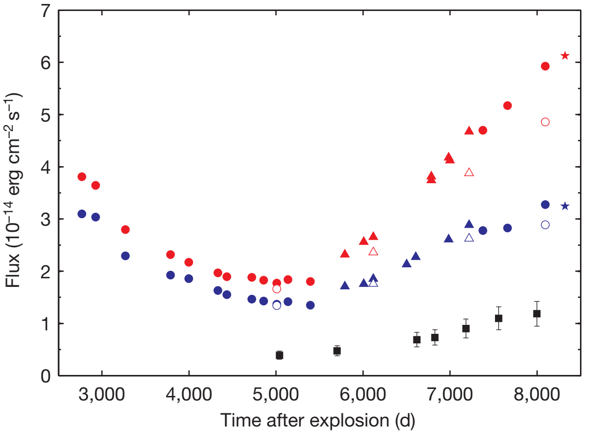


Figure 2 | Light curves of the ejecta in different wavebands. Data points from the HST Wide Field Planetary Camera 2 (WFPC2), Advanced Camera for Surveys (ACS) and Wide Field Camera 3 (WFC3) are shown as dots, triangles and stars, respectively. Red and blue symbols correspond to the R and B bands, respectively. A correction factor has been applied to the ACS and WFC3 fluxes to account for the differences between these instruments and WFPC2. To quantify the contamination from the brightening circumstellar ring, we created detailed models of the ring for a number of different epochs (Supplementary Information, section 1, and Supplementary Figs 3 and 4). The open symbols show the ejecta fluxes after the contribution from the ring has been removed. Although the contamination from the ring increases with time, it never exceeds ~20% of the flux from the ejecta. The statistical errors in the HST fluxes are smaller than the data points, and systematic errors are discussed in the Supplementary Information, section 1. The black squares show the flux of the [Ca II] doublet lines at rest wavelengths 7,291.5 Å and 7,323.9 Å, measured using the Ultraviolet and Visual Echelle Spectrograph of the Very Large Telescope at the European Southern Observatory (error bars, s.d.). These lines are free from contamination from the circumstellar ring and the reverse shock. We note the decreasing flux during the first ~5,000 d and the increase thereafter, indicating an extra source of energy. A further indication that the energy source has changed is that the colour, determined from the flux ratio between the B and R bands, changes from a level of ~0.8 up to day 5,000 to a value close to ~0.6 on day 8,000.

absorbed in the core region of the ejecta (corresponding to speeds less than 5,000 km s^{-1}), where most of the heavy elements reside. At an energy of ~0.35 keV, which corresponds to the temperature of the dominant component in the X-ray spectrum[18], the fraction of flux absorbed by the ejecta at t_{yr} years can be approximated by $1.6\times10^{-3}t_{yr}^{1.67}$ (the increase with time is mainly due to the increasing solid angle of the expanding ejecta, assumed to be spherical, as seen from the ring). This gives a present-day absorbed X-ray luminosity of $\sim5.0\times10^{35}$ erg s^{-1}. In this calculation, we have neglected the weaker, highest-energy component that contributes to the X-ray spectrum[18]. We note that this does not significantly affect the estimate of the absorbed flux, although the hard X-rays may be important owing to their greater penetrating power.

To model the ejecta light curve produced by input from the X-rays, we scaled the observed X-ray flux[6] by the fraction absorbed at 0.35 keV, multiplied the resulting flux by a constant (corresponding to the conversion efficiency from X-rays to optical emission) and added this to the radioactive energy input. Figure 3 shows the scaled X-ray flux together with the observed light curves. This model follows the general trend of the observed fluxes in both bands, although we note that a more accurate model would need to take into account the detailed shape of the X-ray spectrum and the reprocessing of the X-rays into optical emission. The required conversion efficiency from X-rays to optical emission in our model is 5.0% in the R band and 3.1% in the B band.

The conversion of X-rays to optical/infrared emission is similar to that of the ^{44}Ti positrons. Both involve degradation of non-thermal electrons into heating, ionization and excitation. For a typical ionization fraction of 10^{-3}–10^{-2}, the expected efficiency of conversion from X-rays to Hα emission (the dominant line in the R band) is ~5% (Supplementary Information, section 3). This conversion factor is consistent with the scaling factor we used to model the light curve. Similar arguments apply to the B band. Furthermore, the density in the core is high enough for the timescale of recombination to be shorter than the expansion timescale, ensuring a balance between the energy input and output.

Other possible explanations for the increase in flux include input from a central pulsar[19], a transition from optically thick to optically thin dust and positron leakage from the iron-rich regions. We find that input from a pulsar is unlikely for several reasons. In particular, it would be a strange coincidence for the emission from the pulsar to mimic the increasing X-ray flux from the ring interaction. Also, we expect the energy input from a pulsar to be concentrated towards the low-velocity material at the centre of the ejecta, but observations of the Hα and [Ca II] lines show that the increase occurs for speeds up to ~5,000 km s^{-1}. We also note that constraints on a point source at the centre of SN 1987A have already been obtained using HST data

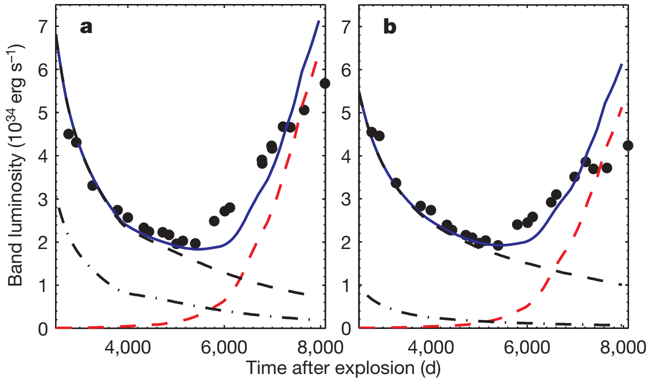


Figure 3 | Evolution of the luminosity from the ejecta in the R and B bands. **a**, R band; **b**, B band. The black dashed lines show a model with only radioactive input, mainly from ^{44}Ti. The ^{44}Ti mass used for the model is $1.4\times10^{-4}M_{\odot}$ (ref. 12), as determined from a detailed analysis that takes into account the effects of scattering and fluorescence of the ultraviolet flux, as well as a correction for internal dust[23,24] (here taken to have a 50% covering factor of the core inside the radius where the ejecta speed is 2,000 km s^{-1}). The dot–dash lines show the light curves with no ^{44}Ti, illustrating the necessity of including this isotope. The red dashed lines show a model based on a constant fraction of the observed X-ray flux[6], corrected for the fraction of flux absorbed by the ejecta (Supplementary Fig. 7). The blue solid line shows the sum of both models (black and red dashed). The R and B bands contain 5% and 3% of the total bolometric flux, respectively, and we expect these fractions to remain roughly constant in time. This is because the relative amount of energy resulting in heating, ionization and excitation will remain nearly constant as long as the ionization by the X-rays is $\lesssim10^{-2}$. It is clear from this figure that there is a transition at ~5,000 d from a radioactivity-dominated phase to a phase dominated by the X-ray input from the collision with the ring.

taken near the minimum of the ejecta light curve[20]. A change in the properties of the dust or a transition in the positron deposition process is also unable to explain the observed increase in flux quantitatively (Supplementary Information, section 4).

We conclude that SN 1987A has made a transition from a radioactively dominated phase to a phase dominated by the X-ray input from the ring collision. This conclusion has interesting implications for the observed morphology. In particular, most of the X-rays are likely to be absorbed at the boundary of the ejecta core, where the density increases rapidly. This may lead to the light from the ejecta being emitted in a ring that is concentrated in the plane of the circumstellar ring. The 'hole' in the ejecta (Fig. 1), which has become more pronounced since about 2001, may in fact be a result of this rather than reflecting the true density distribution or dust obscuration. The asymmetric morphology seen at speeds of $\lesssim$3,000 km s^{-1} in the near-infrared [Si I] and [Fe II] lines[21] is, however, likely to be intrinsic to the metal core. By studying future changes in the morphology of the ejecta, we will be able to understand the origin of this asymmetry

In the future, the density of the ejecta will decrease and the fraction of X-rays absorbed will grow (Supplementary Fig. 7). As a result, the ionization will increase and a smaller fraction of the X-ray flux will produce line excitation. A larger fraction will go into heating, leading to an increase in the mid-infrared flux and a flattening of the optical light curves. In time, the X-rays will also penetrate deeper layers of the ejecta, thereby allowing us to probe the chemical structure of the innermost ejecta. This will be a novel form of X-ray tomography.

Received 31 January; accepted 31 March 2011.
Published online 8 June 2011.

1. Woosley, S. E., Heger, A. & Weaver, T. A. The evolution and explosion of massive stars. *Rev. Mod. Phys.* **74,** 1015–1071 (2002).
2. Fransson, C. & Kozma, C. Radioactivities and nucleosynthesis in SN 1987A. *N. Astron. Rev.* **46,** 487–492 (2002).
3. McKee, C. F. in *Young Supernova Remnants* (eds Holt, S. S. & Hwang, U.) 17–28 (Am. Inst. Phys. Conf. Proc. 565, Springer, 2001).
4. Morris, T. & Podsiadlowski, P. The triple-ring nebula around SN 1987A: fingerprint of a binary merger. *Science* **315,** 1103–1105 (2007).
5. Gröningsson, P. *et al.* Time evolution of the line emission from the inner circumstellar ring of SN 1987A and its hot spots. *Astron. Astrophys.* **492,** 481–491 (2008).
6. Racusin, J. L. *et al.* X-ray evolution of SNR 1987A: the radial expansion. *Astrophys. J.* **703,** 1752–1759 (2009).
7. Zanardo, G. *et al.* Multifrequency radio measurements of supernova 1987A over 22 years. *Astrophys. J.* **710,** 1515–1529 (2010).
8. Dwek, E. *et al.* Five years of mid-infrared evolution of the remnant of SN 1987A: the encounter between the blast wave and the dusty equatorial ring. *Astrophys. J.* **722,** 425–434 (2010).
9. Timmes, F. X., Woosley, S. E., Hartmann, D. H. & Hoffman, R. D. The production of ^{44}Ti and ^{60}Co in supernovae. *Astrophys. J.* **464,** 332–341 (1996).
10. Diehl, R. & Timmes, F. X. Gamma-ray line emission from radioactive isotopes in stars and galaxies. *Publ. Astron. Soc. Pacif.* **110,** 637–659 (1998).
11. Shigeyama, T. & Nomoto, K. Theoretical light curve of SN 1987A and mixing of hydrogen and nickel in the ejecta. *Astrophys. J.* **360,** 242–256 (1990).
12. Jerkstrand, A., Fransson, C. & Kozma, C. The ^{44}Ti-powered spectrum of SN 1987A. *Astron. Astrophys.* (in the press); preprint at ⟨http://arxiv.org/abs/1103.3653⟩ (2011).
13. Michael, E. *et al.* Hubble Space Telescope observations of high-velocity Lyα and Hα emission from supernova remnant 1987A: the structure and development of the reverse shock. *Astrophys. J.* **593,** 809–830 (2003).
14. Smith, N. *et al.* The reverse shock of SNR 1987A at 18 years after outburst. *Astrophys. J.* **635,** L41–L44 (2005).
15. Heng, K. *et al.* Evolution of the reverse shock emission from SNR 1987A. *Astrophys. J.* **644,** 959–970 (2006).
16. France, K. *et al.* Observing supernova 1987A with the refurbished Hubble Space Telescope. *Science* **329,** 1624–1627 (2010).
17. Blinnikov, S., Lundqvist, P., Bartunov, O., Nomoto, K. & Iwamoto, K. Radiation hydrodynamics of SN 1987A. I. Global analysis of the light curve for the first 4 months. *Astrophys. J.* **532,** 1132–1149 (2000).
18. Zhekov, S. A., Park, S., McCray, R., Racusin, J. L. & Burrows, D. N. Evolution of the Chandra CCD spectra of SNR 1987A: probing the reflected-shock picture. *Mon. Not. R. Astron. Soc.* **721,** 518–529 (2010).
19. Woosley, S. E., Hartmann, D. & Pinto, P. A. Hard emission at late times from SN 1987A. *Astrophys. J.* **346,** 395–404 (1989).
20. Graves, G. J. M. *et al.* Limits from the Hubble Space Telescope on a point source in SN 1987A. *Astrophys. J.* **629,** 944–959 (2005).
21. Kjær, K., Leibundgut, B., Fransson, C., Jerkstrand, A. & Spyromilio, J. The 3-D structure of SN 1987A's inner ejecta. *Astron. Astrophys.* **517,** A51–A60 (2010).
22. Chugai, N. N., Chevalier, R. A., Kirshner, R. P. & Challis, P. M. Hubble Space Telescope spectrum of SN 1987A at an age of 8 years: radioactive luminescence of cool gas. *Astrophys. J.* **483,** 925–940 (1997).
23. Lucy, L. B., Danziger, I. J., Gouiffes, C. & Bouchet, P. in *Supernovae* (ed. Woosley, S. E.) 82–94 (Springer, 1991).
24. Wooden, D. H. *et al.* Airborne spectrophotometry of SN 1987A from 1.7 to 12.6 microns: time history of the dust continuum and line emission. *Astrophys. J.* **88** (suppl.), 477–507 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Swedish Research Council and the Swedish National Space Board. Support for the HST observing programme was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc.

Author Contributions J.L. carried out the data reduction and analysis together with G.Ö., P.G., B.L., J.S. and P.C.; C.F. performed the theoretical modelling together with A.J. and C.K.; and J.L. and C.F. wrote the paper. R.P.K. is the principal investigator for the HST/SAINTS collaboration. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.L. (josefin.larsson@astro.su.se) or C.F. (claes@astro.su.se).

LETTER

doi:10.1038/nature10095

Hydrogen-poor superluminous stellar explosions

R. M. Quimby[1], S. R. Kulkarni[1], M. M. Kasliwal[1], A. Gal-Yam[2], I. Arcavi[2], M. Sullivan[3], P. Nugent[4], R. Thomas[4], D. A. Howell[5,6], E. Nakar[7], L. Bildsten[5,8], C. Theissen[9], N. M. Law[1,10], R. Dekany[11], G. Rahmer[11], D. Hale[11], R. Smith[11], E. O. Ofek[11], J. Zolkower[11], V. Velur[11], R. Walters[11], J. Henning[11], K. Bui[11], D. McKenna[11], D. Poznanski[4,12], S. B. Cenko[12] & D. Levitan[13]

Supernovae are stellar explosions driven by gravitational or thermonuclear energy that is observed as electromagnetic radiation emitted over weeks or more[1]. In all known supernovae, this radiation comes from internal energy deposited in the outflowing ejecta by one or more of the following processes: radioactive decay of freshly synthesized elements[2] (typically ^{56}Ni), the explosion shock in the envelope of a supergiant star[3], and interaction between the debris and slowly moving, hydrogen-rich circumstellar material[4]. Here we report observations of a class of luminous supernovae whose properties cannot be explained by any of these processes. The class includes four new supernovae that we have discovered and two previously unexplained events[5,6] (SN 2005ap and SCP 06F6) that we can now identify as members of the same class. These supernovae are all about ten times brighter than most type Ia supernova, do not show any trace of hydrogen, emit significant ultraviolet flux for extended periods of time and have late-time decay rates that are inconsistent with radioactivity. Our data require that the observed radiation be emitted by hydrogen-free material distributed over a large radius (∼10^{15} centimetres) and expanding at high speeds (>10^4 kilometres per second). These long-lived, ultraviolet-luminous events can be observed out to redshifts $z > 4$.

The Palomar Transient Factory[7,8] (PTF) is a project dedicated to finding explosive events and has so far identified over one thousand supernovae. PTF09atu, PTF09cnd and PTF09cwl (also known as SN 2009jh[9]) were detected using the Palomar Observatory's 1.2-m Samuel Oschin Telescope during commissioning of the PTF system, in 2009, and PTF10cwr[10–12] (SN 2010gx[13]) was detected the following year (Fig. 1; see Supplementary Information, section 1). As with other supernova candidates, optical spectra for classification were obtained using the W. M. Keck Observatory's 10-m Keck I telescope, Palomar Observatory's 5.1-m Hale Telescope, and the 4.2-m William Herschel Telescope. The spectra (Fig. 2) show broad absorption dips at short wavelengths and mostly smooth continua at longer wavelengths. We further identify narrow absorption features in the PTF spectra from the Mg II doublet at rest wavelengths 2,796 Å and 2,803 Å, and measure redshifts of $z = 0.501$, 0.258, 0.349 and 0.230 for PTF09atu, PTF09cnd, PTF09cwl and

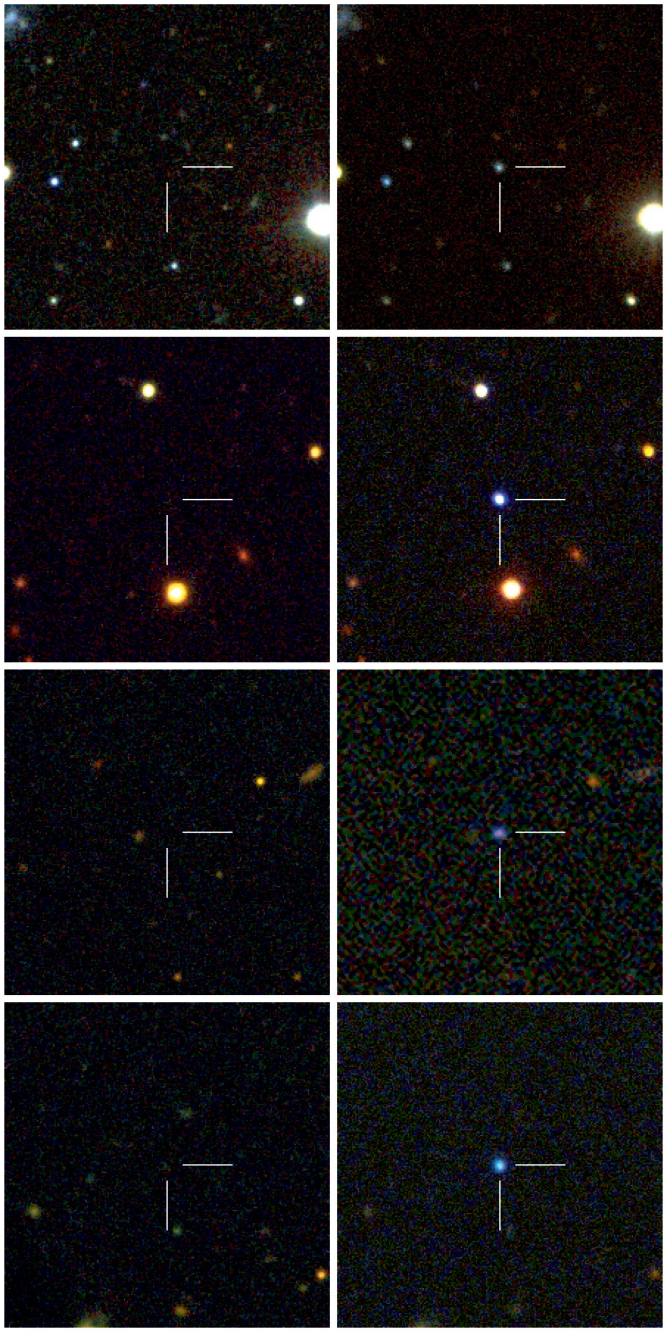


Figure 1 | Ultraviolet-luminous transients discovered by the PTF. Left: before explosion; right: after explosion; top to bottom: PTF09atu, PTF09cnd, PTF09cwl and PTF10cwr. Each tile shows a false-colour image constructed by assigning image data from three separate band passes to red, blue and green (g, r and i bands, respectively, for PTF09atu; u, g or V, and r bands for PTF09cnd, PTF09cwl and PTF10cwr). In each case, Sloan Digital Sky Survey reference data form the pre-explosion image. The post-explosion images are composed from observations made with the Palomar Observatory's 1.5-m telescope, the Wise Observatory's 1.0-m telescope and the Ultraviolet/Optical Telescope on board NASA's Swift satellite.

[1]Cahill Center for Astrophysics 249-17, California Institute of Technology, Pasadena, California 91125, USA. [2]Benoziyo Center for Astrophysics, Faculty of Physics, Weizmann Institute of Science, 76100 Rehovot, Israel. [3]Department of Physics (Astrophysics), University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK. [4]Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. [5]Las Cumbres Observatory Global Telescope Network, 6740 Cortona Drive, Suite 102, Goleta, California 93117, USA. [6]Department of Physics, University of California, Santa Barbara, Broida Hall, Santa Barbara, California 93106, USA. [7]Raymond and Beverly Sackler School of Physics & Astronomy, Tel Aviv University, Tel Aviv 69978, Israel. [8]Kavli Institute for Theoretical Physics, Kohn Hall, University of California, Santa Barbara, California 93106, USA. [9]University of California, San Diego, Department of Physics, 9500 Gilman Drive, La Jolla, California 92093, USA. [10]Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St George Street, Toronto, Ontario M5S 3H4, Canada. [11]Caltech Optical Observatories, California Institute of Technology, Pasadena, California 91125, USA. [12]Astronomy Department, University of California, Berkeley, 601 Campbell Hall, Berkeley, California 94720, USA. [13]Department of Physics, California Institute of Technology, Pasadena, California 91125, USA.

PTF10cwr, respectively. After combining the three available spectra of the SCP 06F6 transient, we find that the data correlate to the PTF sample and may also show narrow Mg II absorption with redshift $z = 1.189$ (Supplementary Information, section 4). Similarly to all PTF events, SN 2005ap ($z = 0.283$) shows a distinct W-shaped absorption feature near rest wavelength 4,300 Å. Although the broad spectral features of SN 2005ap are systematically shifted to higher speeds, the overall resemblance to the other PTF events is striking. The PTF discoveries bridge the redshift gap between SCP 06F6 and SN 2005ap and link these once disparate events, thus unifying them all into a single class.

With the redshifts above and a standard flat cosmology with Hubble parameter $H_0 = 71$ and matter energy density $\Omega_m = 0.27$, the peak absolute u-band AB magnitudes[14] for the PTF transients in the rest frame are near -22 mag and that for SCP 06F6 is near -22.3 mag (Fig. 3). The ~50-d rise of SCP 06F6 to maximum in the rest frame is compatible with the PTF sample, although there seems to be some diversity in the rise and decline timescales. To power these high peak magnitudes with radioactivity, several solar masses ($M_\odot$) of ^{56}Ni are needed ($>10M_\odot$, following ref. 15), and yet in the rest frame V band, the post-maximum decline rates of the PTF events are all >0.03 mag d^{-1}, which is a few times higher than the decay rate of ^{56}Co (the long-lived daughter nucleus of ^{56}Ni). These are therefore not radioactively powered events.

Next we check whether the observed photons could have been deposited by the explosion shock as it traversed the progenitor star. The photospheric radius, R_{ph}, that we infer for PTF09cnd at peak luminosity, on the basis of the observed temperature and assuming blackbody emission, is $R_{ph} \approx 5 \times 10^{15}$ cm (Supplementary Fig. 1). If the radiated photons were generated during the star explosion, then adiabatic losses would result in only a fraction R_*/R_{ph} of the energy remaining in the radiation at any given time, where R_* is the initial stellar radius. Given that the energy radiated around the peak is $\sim10^{51}$ erg and that $R_*/R_{ph} < 10^{-3}$ for virtually any hydrogen-stripped progenitor, this model requires an unrealistic total explosion energy of $>10^{54}$ erg. In fact, the large radius and the duration of PTF09cnd leave almost no place for adiabatic losses (Supplementary Information, section 5), implying that the internal energy must have been deposited at a radius that is not much smaller than R_{ph}.

Integrating the rest frame g-band light curve and assuming no bolometric correction, we find that PTF09cnd radiated $\sim1.2 \times 10^{51}$ erg. A similar analysis of the SCP 06F6 data gives a radiated energy of $\sim1.7 \times 10^{51}$ erg. We also fit Planck functions to the ultraviolet and optical observations of PTF09cnd (Supplementary Fig. 1) and find an approximate bolometric output of $\sim1.7 \times 10^{51}$ erg. The derived blackbody radii indicate a photospheric expansion speed of $v_{ph} \approx 14{,}000$ km s^{-1}. If the main source of luminance were the conversion of kinetic energy, then the bolometric energy would require $\sim1M_\odot$ of material at this speed, assuming a conversion efficiency of 100%. A more realistic efficiency factor would make the minimum mass a few times larger. Because no traces of hydrogen are seen in any of the spectra (Supplementary Information, section 3), interaction with ordinary hydrogen-rich circumstellar material is ruled out. We thus conclude that these events cannot be powered by any of the commonly invoked processes driving known supernova classes.

The early spectra presented here are dominated by oxygen lines and do not show calcium lines, iron lines or other features commonly seen in ordinary core-collapse supernovae. The lack of metals is particularly noticeable in the ultraviolet flux, which is typically depleted by absorption. These events are hosted by low-luminosity galaxies that may provide a subsolar progenitor environment (Supplementary Information, section 6). The new class of events we have identified is thus observationally characterized by extreme peak luminosities, short decay times inconsistent with radioactivity, and very hot early spectra with significant ultraviolet flux and lacking absorption lines from heavy elements such calcium and iron, which are commonly seen in all other types of supernova.

These observations require a late deposition of a large amount of energy ($>10^{51}$ erg) into hydrogen-poor, rapidly expanding material (slow-moving material would produce narrow spectroscopic features, which are not observed). We point out two possible physical processes that can perhaps power these superluminous sources. One is a strong interaction with a massive, rapidly expanding, hydrogen-free shell.

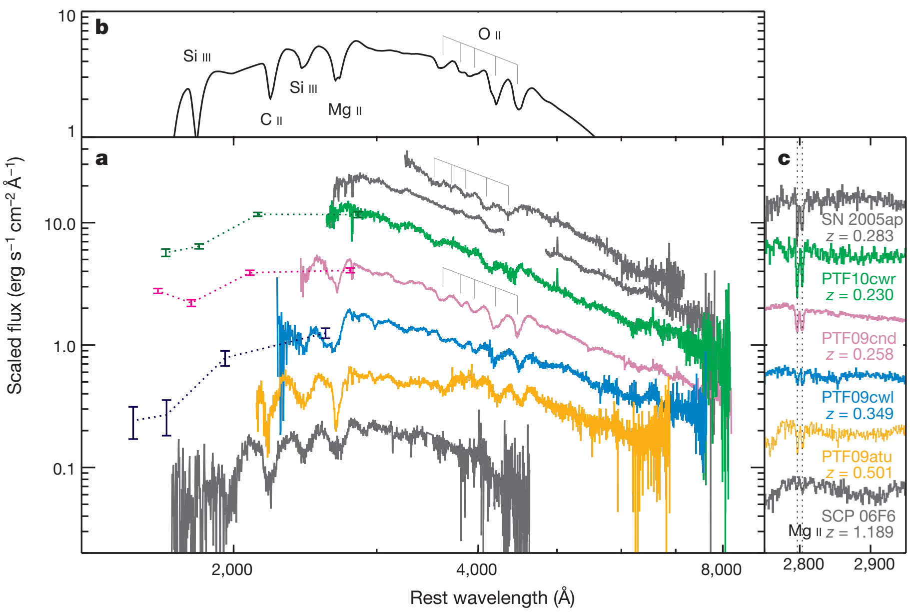


Figure 2 | Spectral energy distributions of the SN 2005ap-like sample. **a**, Top to bottom: SN 2005ap on 2005 March 7 and again on March 16 (grey), PTF10cwr on 2010 March 18 (green), PTF09cnd on 2009 August 25 (purple), PTF09cwl on 2009 August 25 (blue), PTF09atu on 2009 July 20 (orange) and an average of all three SCP 06F6 spectra presented in ref. 6 (grey). The spectra have been adjusted to remove the effect of redshift, binned and scaled arbitrarily for display purposes. Broadband flux densities from the Swift observations (scaled to join the spectra in the u band) are plotted for PTF09cnd, PTF09cwl and PTF10cwr with 1σ error bars. Five absorption bands are marked by the combs above SN 2005ap and PTF09cnd, the former being ~7,000 km s^{-1} faster. **b**, These features can be well fitted by O II using the highly parametric spectral synthesis code SYNOW[21] (Supplementary Information, section 2). SYNOW fits also suggest that C II and Mg II can account for the respective features at 2,200 Å and 2,700 Å. The fit to the line at 2,500 Å is improved with the addition of Si III. The model shown has a photospheric speed of 15,000 km s^{-1}. **c**, Close-up views of the narrow Mg II doublet, from which we derive the redshifts.

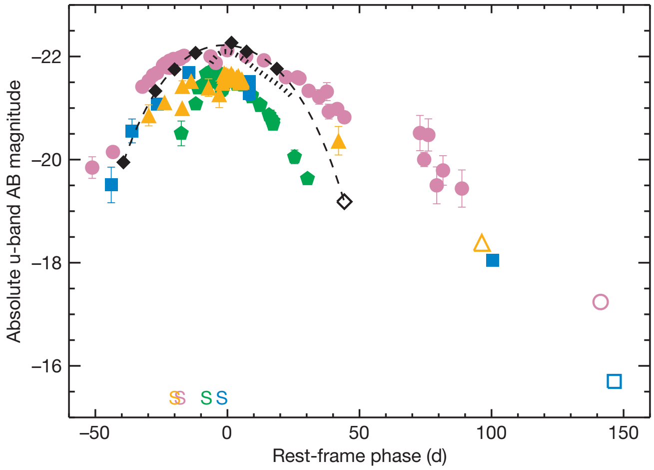


Figure 3 | Luminosity evolution of the SN 2005ap-like sample. Shown are the SCP 06F6 transient (diamonds), SN 2005ap (hatching), PTF09atu (orange triangles), PTF09cnd (purple dots), PTF09cwl (blue squares) and PTF10cwr (green pentagons). In each case, we transform the observed photometry to absolute u-band magnitudes by correcting both for distance and differences in the effective rest-frame band pass introduced by the redshifts. For SCP 06F6, the observed i band is similar to the rest frame u band, so the correction factor is nearly independent of the spectral properties[22]. For the PTF sample, however, the correction factor varies over time as the supernovae cool. We interpolated the observed spectra of PTF09cnd to phases appropriate for the B-, g-, V- and r-band observations of the PTF sample to calculate the correction factors. Errors bars representing 1σ (excluding the colour correction) are shown when larger than the plotting symbols. The colour corrections for PTF09atu and PTF09cwl near day 100 are very uncertain ($\sim$0.5 mag). We have not removed possible host light contaminating the late-time observations of PTF09atu, PTF09cnd and PTF09cwl (open symbols). Thus, these measurements represent upper limits on the supernova light. Host galaxy light may in fact dominate the final PTF09cwl observation. The final observation of SCP 06F6 (open diamond) is a 2.5σ detection made from the ground. Along the abscissa, with appropriately colour-coded 'S's we note the phases of spectra shown in Fig. 2.

Such a situation is naturally produced by extremely massive stars with initial masses in the range $90\ M_\odot \lesssim M_\mathrm{i} \lesssim 130\ M_\odot$, which are expected[16,17] to undergo violent pulsations—perhaps driven by the pair instability—that strip their outer layers and expel massive, hydrogen-poor shells. The star eventually dies by becoming a stripped-envelope, core-collapse supernova, which may interact with previously ejected carbon- or oxygen-rich shells to drive the observed luminosity[18]. Alternatively, the power source can be a prolonged energy injection by a central engine. For example, a spinning-down nascent magnetar[19,20] can account for the peak luminosities ($>10^{44}\ \mathrm{erg\,s^{-1}}$) and time to peak light (30–50 d) observed for these events, assuming a magnetic field $B \approx (1\text{–}3) \times 10^{14}$ G and a natal spin period of 1–3 ms.

The high luminosities, exceptionally blue spectral energy distributions and volumetric event rates (Supplementary Information, section 7) of this new class of supernovae make them prime targets for high-redshift studies (Supplementary Fig. 3). Because they remain at around maximum luminosity for months to years in the observer frame, these events provide a steady light source to illuminate their environs and any intervening clouds of gas and dust. This creates new opportunities for high-resolution spectroscopy to probe distant star-forming regions in primitive galaxies without the need for rapid scheduling and with the benefit that the luminous supernova beacon eventually fades, allowing the study of the galaxy itself. Indeed, these events promise to be a rich source of results for future 30-m-class-telescope science.

Received 17 February; accepted 1 April 2011.
Published online 8 June 2011.

1. Woosley, S. E. & Weaver, T. A. The physics of supernova explosions. *Annu. Rev. Astron. Astrophys.* **24,** 205–253 (1986).
2. Arnett, W. D. Type I supernovae. I - Analytic solutions for the early part of the light curve. *Astrophys. J.* **253,** 785–797 (1982).
3. Grassberg, E. K., Imshennik, V. S. & Nadyozhin, D. K. On the theory of the light curves of supernovae. *Astrophys. Space Sci.* **10,** 28–51 (1971).
4. Chevalier, R. A. Self-similar solutions for the interaction of stellar ejecta with an external medium. *Astrophys. J.* **258,** 790–797 (1982).
5. Quimby, R. M. *et al.* SN 2005ap: a most brilliant explosion. *Astrophys. J.* **668,** L99–L102 (2007).
6. Barbary, K. *et al.* Discovery of an unusual optical transient with the Hubble Space Telescope. *Astrophys. J.* **690,** 1358–1362 (2009).
7. Law, N. M. *et al.* The Palomar Transient Factory: system overview, performance, and first results. *Publ. Astron. Soc. Pacif.* **121,** 1395–1408 (2009).
8. Rau, A. *et al.* Exploring the optical transient sky with the Palomar Transient Factory. *Publ. Astron. Soc. Pacif.* **121,** 1334–1351 (2009).
9. Drake, A. J. *et al.* Supernova 2009jh. *Centr. Bur. Electron. Telegr.* 1958 (2009).
10. Mahabal, A. A. *et al.* Supernova candidates and classifications from CRTS. *Astronomer's Telegr.* 2490 (2010).
11. Quimby, R. M. *et al.* Discovery of a luminous supernova, PTF10cwr. *Astronomer's Telegr.* 2492 (2010).
12. Pastorello, A. *et al.* Detection of PTF10cwr/CSS100313 on PS1 sky survey images and host galaxy identification. *Astronomer's Telegr.* **2504** (2010).
13. Pastorello, A. *et al.* Ultrabright optical transients are linked with type Ic supernovae. *Astrophys. J.* **724,** L16–L21 (2010).
14. Oke, J. B. & Gunn, J. E. Secondary standard stars for absolute spectrophotometry. *Astrophys. J.* **266,** 713–717 (1983).
15. Gal-Yam, A. *et al.* Supernova 2007bi as a pair-instability explosion. *Nature* **462,** 624–627 (2009).
16. Woosley, S. E., Blinnikov, S. & Heger, A. Pulsational pair instability as an explanation for the most luminous supernovae. *Nature* **450,** 390–392 (2007).
17. Umeda, H. & Nomoto, K. How much ^{56}Ni can be produced in core-collapse supernovae? Evolution and explosions of 30–100 $M_\odot$ stars. *Astrophys. J.* **673,** 1014–1022 (2008).
18. Chevalier, R. A. & Irwin, C. M. Shock breakout in dense mass loss: luminous supernovae. *Astrophys. J.* **729,** L6–L9 (2011).
19. Woosley, S. E. Bright supernovae from magnetar birth. *Astrophys. J.* **719,** L204–L207 (2010).
20. Kasen, D. & Bildsten, L. Supernova light curves powered by young magnetars. *Astrophys. J.* **717,** 245–249 (2010).
21. Jeffery, D. J. & Branch, D. in *Supernovae, Jerusalem Winter School for Theoretical Physics* (eds Wheeler, J. C., Piran, T. & Weinberg, S.) 149–247 (World Scientific, 1990).
22. Hogg, D. W., Baldry, I. K., Blanton, M. R. & Eisenstein, D. J. The K correction. Preprint at ⟨http://arxiv.org/abs/astro-ph/0210394⟩ (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Observations were obtained with the Samuel Oschin Telescope and the 60-inch telescope at Palomar Observatory as part of the PTF project, a scientific collaboration between the California Institute of Technology, Columbia University, Las Cumbres Observatory, the Lawrence Berkeley National Laboratory, the National Energy Research Scientific Computing Center, the University of Oxford and the Weizmann Institute of Science. Some of the data presented here were obtained at the W. M. Keck Observatory and the William Herschel Telescope. The National Energy Research Scientific Computing Center, which is supported by the Office of Science of the US Department of Energy, provided staff, computational resources and data storage for this project. Observations by the Weizmann PTF partnership and from the Wise Observatory are supported by grants from the Israel Science Foundation and the US–Israel Binational Science Foundation. We acknowledge support from the US Department of Energy Scientific Discovery through Advanced Computing programme, the Hale Fellowship from the Gordon and Betty Moore foundation, the Bengier Foundation, the Richard and Rhoda Goldman Fund, and the Royal Society.

Author Contributions R.M.Q. initiated, coordinated and managed the project, carried out photometric and spectroscopic observations and analysis, and wrote the manuscript. S.R.K. is the PTF principal investigator and contributed to manuscript preparation. M.M.K. obtained spectroscopy from the Keck I telescope and helped with the P60 observations. A.G.-Y. oversaw the Wise observations and contributed to analysis and manuscript writing. I.A. extracted the Wise photometry and helped obtain Keck I spectra. M.S. carried out and analysed spectroscopic observations from the William Herschel Telescope. P.N. designed and implemented the image-subtraction pipeline that detected the PTF events. R.T. analysed the combined spectra using his automated SYNOW code. D.A.H. helped to identify the PTF spectra as being like that of SN 2005ap. E.N. contributed to the physical interpretation and manuscript writing. L.B. advised during the preparation of the manuscript. C.T. helped vet potential candidates and first identified PTF09atu and PTF09cwl. N.M.L. is the PTF project scientist and oversaw the PTF system. R.D., G.R., D.H., R.S., E.O.O., J.Z., V.V., R.W., J.H., K.B. and D.M. helped to build and commission the PTF system. D.P., S.B.C. and D.L. helped to vet PTF candidates and obtain spectroscopic observations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.M.Q. (quimby@astro.caltech.edu).

doi:10.1038/nature10112

MicroRNAs 103 and 107 regulate insulin sensitivity

Mirko Trajkovski[1,2], Jean Hausser[2,3], Jürgen Soutschek[4], Bal Bhat[4], Akinc Akin[5], Mihaela Zavolan[3], Markus H. Heim[2,6] & Markus Stoffel[1,2]

Defects in insulin signalling are among the most common and earliest defects that predispose an individual to the development of type 2 diabetes[1–3]. MicroRNAs have been identified as a new class of regulatory molecules that influence many biological functions, including metabolism[4,5]. However, the direct regulation of insulin sensitivity by microRNAs *in vivo* has not been demonstrated. Here we show that the expression of microRNAs 103 and 107 (miR-103/107) is upregulated in obese mice. Silencing of miR-103/107 leads to improved glucose homeostasis and insulin sensitivity. In contrast, gain of miR-103/107 function in either liver or fat is sufficient to induce impaired glucose homeostasis. We identify caveolin-1, a critical regulator of the insulin receptor, as a direct target gene of miR-103/107. We demonstrate that caveolin-1 is upregulated upon miR-103/107 inactivation in adipocytes and that this is concomitant with stabilization of the insulin receptor, enhanced insulin signalling, decreased adipocyte size and enhanced insulin-stimulated glucose uptake. These findings demonstrate the central importance of miR-103/107 to insulin sensitivity and identify a new target for the treatment of type 2 diabetes and obesity.

To identify microRNAs (miRNAs) that are deregulated in obesity and insulin resistance, we performed miRNA microarray analysis on the livers of two types of obese mice: *ob/ob* mice and diet-induced-obese (DIO) C57BL/6J mice (Supplementary Table 1a–c). The miR-103/miR-107 family was among the five most-upregulated miRNAs in the livers of both obese models, and the expression of these miRNAs was also reportedly increased in diabetic Goto-Kakizaki rats[6]. The expression levels were validated by northern blotting, demonstrating a twofold to threefold upregulation in the livers of both models (Fig. 1a). The sequences of mature miR-103 and miR-107 differ by one nucleotide at position 21 and cannot be discriminated by northern blotting. By real-time PCR, we could distinguish miR-103 and miR-107 and show that both miRNAs are upregulated in the livers of *ob/ob* and DIO mice (Supplementary Fig. 1a–d). We also measured the expression of these miRNAs in liver biopsies from a cohort of human patients. miR-103 and miR-107 levels were similar in normal subjects and in subjects with viral hepatitis, but increased in alcoholic liver disease (ALD), non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH), conditions often associated with diabetes[7]. Furthermore, there was a positive correlation between the subjects' homeostatic model assessment (HOMA) index and miR103/107 expression levels (Fig. 1b and Supplementary Fig. 2a–c), indicating an association of these miRNAs with insulin resistance.

To investigate the effect of elevated miR-103/107 expression, we generated recombinant adenovirus expressing miR-107 (ad-107/GFP). Injection of wild-type mice with ad-107/GFP (Supplementary Fig. 3a, b) caused a rise in both random and fasting blood-glucose levels, and also in insulin levels (Fig. 1c, d). It impaired glucose tolerance after an intra-peritoneal glucose injection and decreased insulin sensitivity relative to that in control ad-GFP-infected mice (Fig. 1e, f). Hepatic overexpression of miR-107 resulted in increased glucose production during an intraperitoneal pyruvate-tolerance test (Fig. 1g). The increase in hepatic glucose production was accompanied by augmented expression of glucose 6-phosphatase, phosphoenolpyruvate carboxykinase, pyruvate carboxylase and fructose 1,6-biphosphatase, indicating that increased gluconeogenesis is the primary cause of the elevated glucose levels (Fig. 1h). These data show that gain-of-function of miR-107 in the liver decreases insulin sensitivity and enhances hepatic glucose production.

To study the effect of miR-103/107 silencing, we first tested whether antagomirs[8] would inhibit both miR-103 and miR-107. Northern blot analysis of miR-103 and miR-107 showed that antagomir-103 (ant-103) effectively and specifically silenced both miRNAs in liver and fat (Supplementary Fig. 3c, d). Markers for liver damage and inflammation were unaffected by the treatment (data not shown). Application of ant-103 did not affect blood glucose levels in chow-fed wild-type mice but it did lower plasma glucose levels in *ob/ob* mice when compared to mice treated with PBS or with the controls ant-124 (scrambled) or mismatch-ant-103 (ant-MM103) (Fig. 2a, b). Similar effects were observed in DIO mice (Fig. 2c). Glucose-tolerance and insulin-tolerance tests showed that there was improved glucose tolerance and insulin sensitivity in both *ob/ob* and DIO mice that were injected with ant-103 (Fig. 2d–f). A pyruvate-tolerance test revealed that *de novo* hepatic glucose production was reduced (Fig. 2g) and this finding was supported by a reduction in hepatic expression of glucose 6-phosphatase, pyruvate carboxylase and fructose 1,6-biphosphatase in ant-103-treated mice (Fig. 2h). In addition, liver glycogen content was increased and plasma insulin levels were decreased in ant-103-treated *ob/ob* and DIO animals (Fig. 2i, j). Metabolic and energy-expenditure studies carried out in metabolic cages showed that *ob/ob* mice lacking miR-103 expression had increased O_2 consumption and CO_2 production, as well as moderately elevated body temperature, but that their food intake was similar to control mice. Gene expression analysis in adipocytes from ant-103-treated mice revealed increased levels of β-oxidation genes (carnitine palmitoyltransferase 1a (*Cpt1a*), peroxisomal acyl-coenzyme A oxidase 3 (*Acox3*) and very long chain acyl-coenzyme A dehydrogenase (*Acadvl*)) but there were no changes in the lipogenic genes acetyl-coenzyme A carboxylase alpha and beta (*Acaca* and *Acacb*, also known as *Acc1* and *Acc2*) (Supplementary Fig. 4a–d). Two independent indicators of insulin sensitivity, the glucose infusion rate and clamp glucose turnover, were improved during hyperinsulinaemic-euglycaemic clamp studies in ant-103-treated *ob/ob* mice compared to PBS-injected controls. Hepatic glucose production was decreased and glucose uptake in adipose tissue was enhanced in ant-103-treated animals. In contrast, treatment with ant-103 did not improve glucose uptake in skeletal muscle (Supplementary Table 2). Together, these data demonstrate that silencing of miR-103/107 enhances insulin sensitivity in liver and adipose tissue.

To test the contribution of the liver to the effect on insulin sensitivity, we delivered ant-103 and control ant-MM103 specifically to the liver through liposomal formulations. Liposomal ant-103 induced the specific silencing of miR-103 in the liver, but not in fat and muscle. Silencing of miR-103/107 in the livers of *ob/ob* mice had no effect on

[1]Institute for Molecular Systems Biology, ETH Zurich, Wolfgang-Pauli Strasse 16, CH-8093 Zurich, Switzerland. [2]Competence Center of Systems Physiology and Metabolic Disease, ETH Zurich, Schafmattstrasse 18, HPM F 39.1 CH-8093 Zurich, Switzerland. [3]Biozentrum Basel, University of Basel, Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland. [4]Regulus Therapeutics Inc., 3545 John Hopkins Court, San Diego, California 92121-1121, USA. [5]Alnylam Pharmaceuticals, 300 Third Street, Cambridge, Massachusetts 02142, USA. [6]University Hospital Basel, Hebelstrasse 20, CH-4031 Basel, Switzerland.

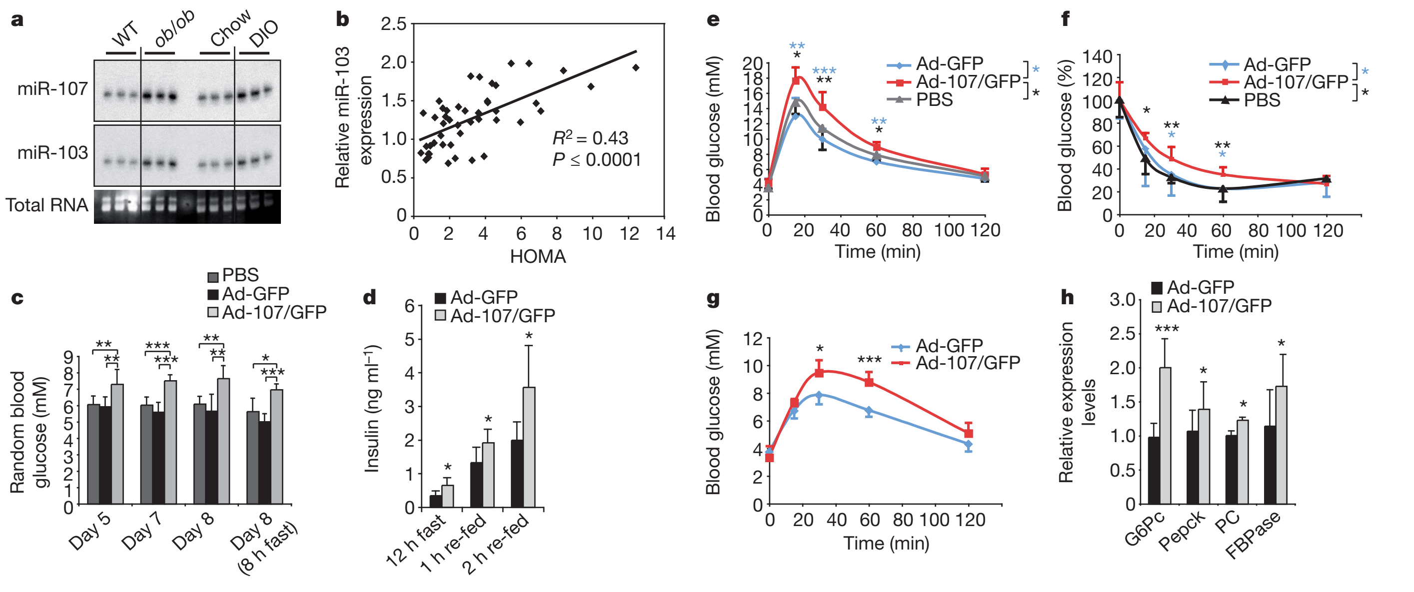


Figure 1 | Hepatic overexpression of miR-107 induces hyperglycaemia. **a**, Northern blot of liver RNA from C57BL/6J (WT), *ob/ob*, chow-fed or DIO mice, as indicated ($n = 3$). The loading control labelled 'total RNA' is stained with ethidium bromide. **b**, Correlation between relative miR-103 levels (Rel. miR-103) and HOMA index in a group of humans including healthy individuals ($n = 6$), chronic hepatitis B and hepatitis C virus-infected individuals (HBV, $n = 6$; HCV, $n = 7$), patients with alcoholic steatohepatitis ($n = 4$), patients with non-alcoholic fatty liver disease ($n = 13$) and patients with non-alcoholic steatohepatitis ($n = 13$). **c**, Blood glucose levels of C57BL/6J mice injected with PBS, ad-GFP or ad-107/GFPBL/ ($n = 6$). **d**, Plasma insulin levels of C57BL/6J mice treated as in **c**, after a 12 h fast followed by re-feeding. **e**–**g**, Glucose-tolerance test (**e**), insulin-tolerance test (**f**) and pyruvate-tolerance test (**g**) in mice injected with ad-GFP or ad-107/GFP. **h**, Relative mRNA expression of genes encoding glucose 6-phosphatase (G6Pc), phosphoenol-pyruvate carboxykinase (Pepck), pyruvate carboxylase (PC) and fructose 1,6-biphosphatase (FBPase) from livers of mice as in **e**–**g**. Expression is normalized to the *36B4* gene, encoding the acidic ribosomal phosphoprotein P0 (RPLP0) ($n = 5$). Means ± s.d. are shown for all panels. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

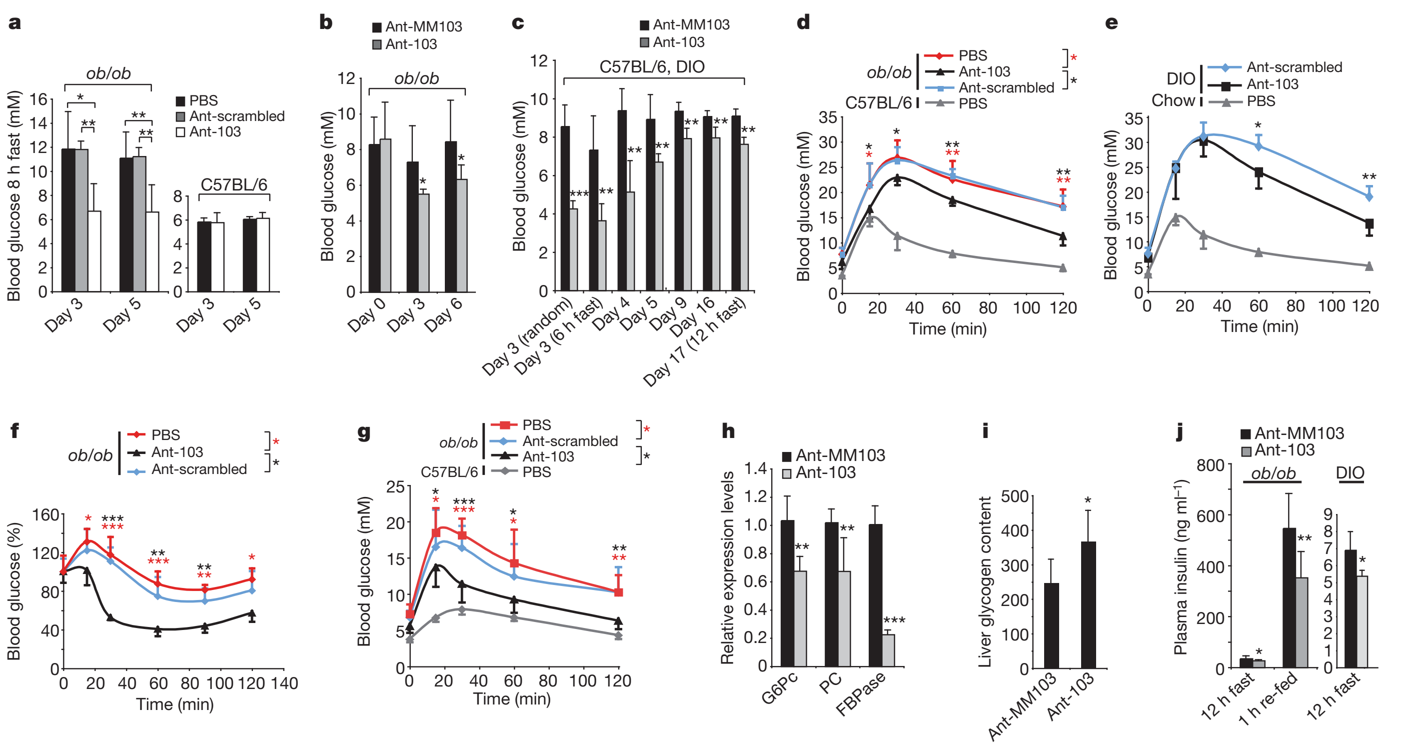


Figure 2 | Silencing of miR-103 and miR-107 alleviates hyperglycaemia in diabetic mice. **a**–**c**, Blood glucose levels of ant-103-treated or control-treated (PBS, scrambled or ant-MM103) *ob/ob* ($n = 6$), C57BL/6J ($n = 5$) or DIO ($n = 5$) mice. Days after treatment and random or fasting conditions are noted for each measurement. **d**, **e**, Glucose-tolerance tests in control or ant-103-injected *ob/ob* (**d**) or DIO (**e**) mice compared to control C57BL/6J or chow-fed mice ($n = 5$). **f**, **g**, Insulin-tolerance test (**f**) or pyruvate-tolerance test (**g**) in control or ant-103-injected *ob/ob* mice ($n = 5$). In **f**, values at the zero time point are normalized to 100%. **h**, Relative mRNA expression of the genes for G6Pc, PC and FBPase in livers of *ob/ob* mice 16 d after injection with ant-MM103 or ant-103 ($n = 5$). **i**, Liver glycogen content in *ob/ob* mice 16 d after injection with ant-MM103 or ant-103 ($n = 5$). **j**, Insulin levels in *ob/ob* mice (left, $n = 10$) or DIO mice (right, $n = 5$) 10 d after injection of ant-MM103 or ant-103. In **d**, **e** and **g**, PBS-injected C57BL/6J mice ($n = 5$) are shown as controls. Means ± s.d. are shown for all panels. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

blood glucose levels, plasma insulin levels or glucose-tolerance, insulin-tolerance and pyruvate-tolerance tests (Supplementary Fig. 5a–f), indicating that silencing of miR-103/107 in the liver is not sufficient to reverse insulin resistance in obese mice. Because the expression of miR-103 is about eightfold higher in adipose tissue than in liver and muscle, we examined the effect of miR-103/107 silencing in adipose tissue. Obese (*ob/ob)* mice showed a slight reduction in body weight when miR-103/107 were systemically silenced (Supplementary Fig. 6a). In contrast, specific manipulation of hepatic miR-103/107 expression using liposomal ant-103 or Ad-107/GFP did not affect body weight when compared to that of control-treated mice (data not shown). We therefore used computer tomography to investigate the fat distribution of DIO and *ob/ob* animals after miR-103 silencing. Both DIO and *ob/ob* mice treated with ant-103 showed reduced levels of total fat, owing to a decrease in both subcutaneous and visceral adipose tissue (Fig. 3a, b). Furthermore, organ measurements revealed a decrease in inguinal fat-pad weights in the ant-103-treated group but no weight differences in other organs (Supplementary Fig. 6b). To investigate whether this reduction was due to lower cell numbers or smaller adipocytes, we quantified the mean size of adipocytes from fat tissue sections using automated image-analysis software. Ant-103-treated DIO and *ob/ob* animals had smaller adipocytes than ant-MM103-injected controls (Fig. 3c, d), owing to an increased number of small adipocytes and a decreased number of large ones (Supplementary Fig. 6c–f). A comparison between the decrease in fat-pad size, measured by computer tomography, and the average decrease in adipocyte size showed that ant-103-treated mice had approximately 10–20% more adipocytes than ant-MM103-treated controls.

Because miR-103 has been implicated in adipocyte differentiation[9–13], we explored whether the increase in adipocyte number in miR-103-depleted mice could be attributed to changes in pre-adipocyte differentiation. We induced adipocyte differentiation of isolated stromal-vascular cells from both visceral and subcutaneous fat in the presence of either ant-103 or ant-MM103. Quantification of mature adipocyte numbers by high-content imaging after 8 d in culture demonstrated a 2-fold and 2.5-fold increase in the number of differentiated adipocytes in the ant-103-treated stromal-vascular cells derived from visceral and subcutaneous fat, respectively. This indicates that the absence of miR-103 enhances adipocyte differentiation in a cell-autonomous fashion. Conversely, overexpression of miR-107 decreased the number of differentiated adipocytes (Fig. 3e). The negative effect of miR-103 on pre-adipocyte differentiation was further corroborated by gene expression analysis of an early marker of adipocyte differentiation, CCAAT/enhancer binding protein (C/EBP)-β (*Cebpb*), and of two late markers of adipocyte differentiation, peroxisome proliferator activated receptor-γ (*Pparg*) and adipocyte fatty acid binding protein 4 (*Fabp4*) (Fig. 3f–h).

To test whether fat-specific overexpression of miR-107 affects insulin sensitivity, we injected either ad-107/GFP or ad-GFP into the inguinal fat pads of wild-type mice. The relative size of the inguinal fat deposit was 20.7% ± 6.3% of total body fat, as determined by computer tomography. Expression of miR-107 was increased by about 1.6-fold and was restricted to the fat pad. Levels of blood glucose and insulin in mice injected with ad-107/GFP were increased and glucose-tolerance and insulin-tolerance tests showed decreased glucose tolerance and insulin sensitivity, respectively, supported by an increase in the HOMA index. Furthermore, adipocyte size was increased in the fat pads injected with ad-miR-107, compared to the ad-GFP controls (Supplementary Fig. 7a–g). These data show that overexpression of miR-107 in the fat is sufficient to induce insulin resistance and glucose intolerance.

Smaller adipocytes are associated with increased insulin sensitivity in human and rodent models[14]. To explore whether insulin-stimulated glucose uptake in adipocytes was affected by miR-103 silencing, we isolated primary adipocytes from *ob/ob* mice injected with either ant-103 or ant-MM103 and measured insulin-stimulated D-^{14}C-glucose

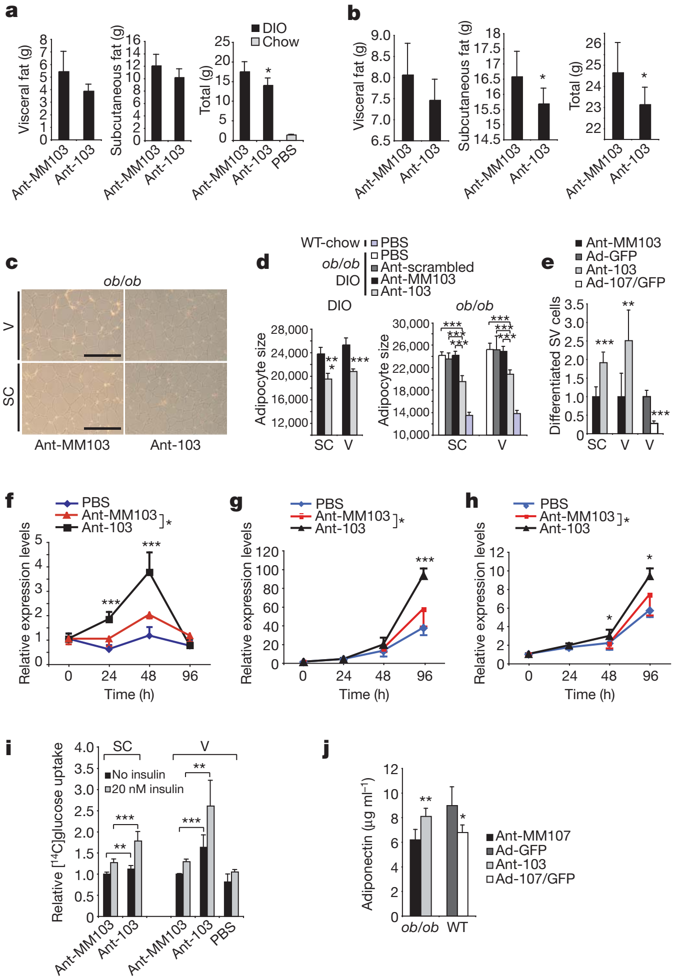


Figure 3 | Silencing of miR-103 decreases total fat by reducing adipocyte size. a, **b**, Fat-pad weights from subcutaneous and visceral adipose tissue of mice injected with ant-MM103 or ant-103. DIO mice (**a**, $n = 4 + 5$, respectively) or *ob/ob* mice (**b**, $n = 6$) were assessed by computer tomography 13 d after injection. **c**, Haematoxylin staining of paraffin sections from subcutaneous (SC) and visceral (V) fat of *ob/ob* mice injected with ant-MM103 or ant-103. Scale bar: 250 µm. **d**, Automated quantification of the average adipocyte size from mice treated as in **c**, with PBS and with ant-scrambled controls. **e**, Automated quantification of differentiated adipocytes after 8 days of differentiation in the presence of ant-103, ant-MM103, ad-GFP or ad-107/GFP. The values shown are normalized to 1. **f**–**h**, Relative mRNA levels (Rel. expression) of *Cebph* (**f**), *Fabp4* (**g**) or *Pparg* (**h**) in stromal-vascular cells differentiated in the presence of PBS, ant-MM103 or ant-103, at the indicated time points. The mRNA expression levels at time point 0 were determined before the antagomir/differentiation treatment. **i**, [^{14}C]Glucose uptake in primary adipocytes isolated from subcutaneous or visceral fat of *ob/ob* mice injected with PBS, ant-MM103 or ant-103. Uptake is normalized to cell numbers. **j**, Adiponectin levels in ant-MM103-injected or ant-103-injected *ob/ob* mice (left, $n = 7$), and in ad-GFP-injected or ad-107/GFP-injected C57BL/6J mice (right, $n = 9$). Means ± s.d. are shown for all panels. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

uptake *in vitro*. Basal and insulin-stimulated glucose uptake was increased in adipocytes from both subcutaneous and visceral fat of ant-103-injected animals (Fig. 3i). Furthermore, adiponectin levels, which correlate positively with insulin sensitivity[15], were increased in ant-103-injected *ob/ob* mice (Fig. 3j). Together, these data show that silencing of miR-103/107 increases insulin sensitivity in adipocytes.

To address the possible mechanism by which miR-103 and miR-107 regulate insulin sensitivity, we performed genome-wide expression analysis using Affymetrix microarrays, comparing livers from C57BL/6J mice infected with ad-107/GFP or ad-GFP. In animals

infected with ad-107/GFP, mRNAs carrying a seed match to miR-107 in the 3′ untranslated region (3′ UTR) were downregulated when compared to transcripts that lacked a miR-107 seed. The data were confirmed by real-time PCR for a subset of miR-107 target genes (Fig. 4a). Out of over 3,000 genes with a 6-mer seed match in the 3′ UTR, the predicted top 100 targets of miR-107/103 were enriched in membrane-related genes and metabolism genes (Supplementary Fig. 8a, b). The gene encoding caveolin-1 (*Cav1*), a key component of caveolae and a mediator of insulin signalling, was among the miR-103/107 seed-containing genes that were downregulated after overexpression of miR-107 in the liver, and upregulated after its silencing (Fig. 4a and Supplementary Fig. 8c). Notably, miR-103 silencing in the fat resulted in an approximately 3.5-fold upregulation of *Cav1* mRNA levels (Fig. 4b). Murine *Cav1* (m*Cav1*) contains three miR-103 sites, whereas human *CAV1* (h*CAV1*) has two seed motifs in the 3′ UTR (Supplementary Fig. 8d, e). Measurements of luciferase activity in HEK 293 cells transfected with reporter plasmids containing the 3′ UTRs of m*Cav1*or h*CAV1* showed reduced expression of these constructs in the presence of miR-103 (Fig. 4c). By mutating the conserved seed, we could fully reverse the miR-103-induced decrease in luciferase activity in both m*Cav1* and h*CAV1* constructs (Fig. 4c). Overexpression of miR-103 also led to an approximately twofold decrease in endogenous Cav1 levels in HEK 293 cells compared to controls (Fig. 4d). Conversely, ant-103, but not ant-scrambled or PBS controls, increased Cav1 levels in HEK 293 cells (Fig. 4e). Taken together, these data demonstrate that Cav1 is a direct target of miR-103 in both mouse and human cells.

Cav1 is the principal protein of caveolae[16], distinct lipid- and cholesterol-enriched vascular invaginations at the plasma membrane. Cav1 activates insulin signalling, probably by stabilizing caveolae and their associated insulin receptors[17]. Peptides corresponding to the scaffolding domain of Cav1 and Cav3 potently stimulate insulin-receptor-kinase activity[18]. Furthermore, overexpression of Cav3 augments insulin-stimulated phosphorylation of insulin receptor substrate 1 (ref. 18) and increases hepatic insulin-receptor phosphorylation in response to insulin stimulation, thereby improving the overall glucose metabolism of diabetic mice[19]. *Cav1*-null mice are phenotypically normal on a chow diet but develop insulin resistance on a high-fat diet owing to decreased insulin-receptor expression and diminished insulin-receptor signalling in adipose tissue[20]. We investigated whether insulin signalling correlated with miR-103/107-mediated changes in Cav1 expression. In the fat and liver of *ob/ob* mice, silencing of miR-103/107 resulted in increased Cav1 levels, whereas no expression of this protein could be detected in skeletal muscle (Fig. 4f–h). The expression of insulin receptor β-subunit (IRβ) in adipocytes was increased and insulin-stimulated levels of phosphorylated Akt1 and IRβ (pAkt1 and pIRβ) were augmented in the fat and liver of ant-103-treated mice (Fig. 4f, g). In contrast, insulin signalling was not enhanced in the skeletal muscle of ant-103-treated mice (Fig. 4h). In addition, wild-type mice in which ad-miR-107 was injected into the inguinal fat pad

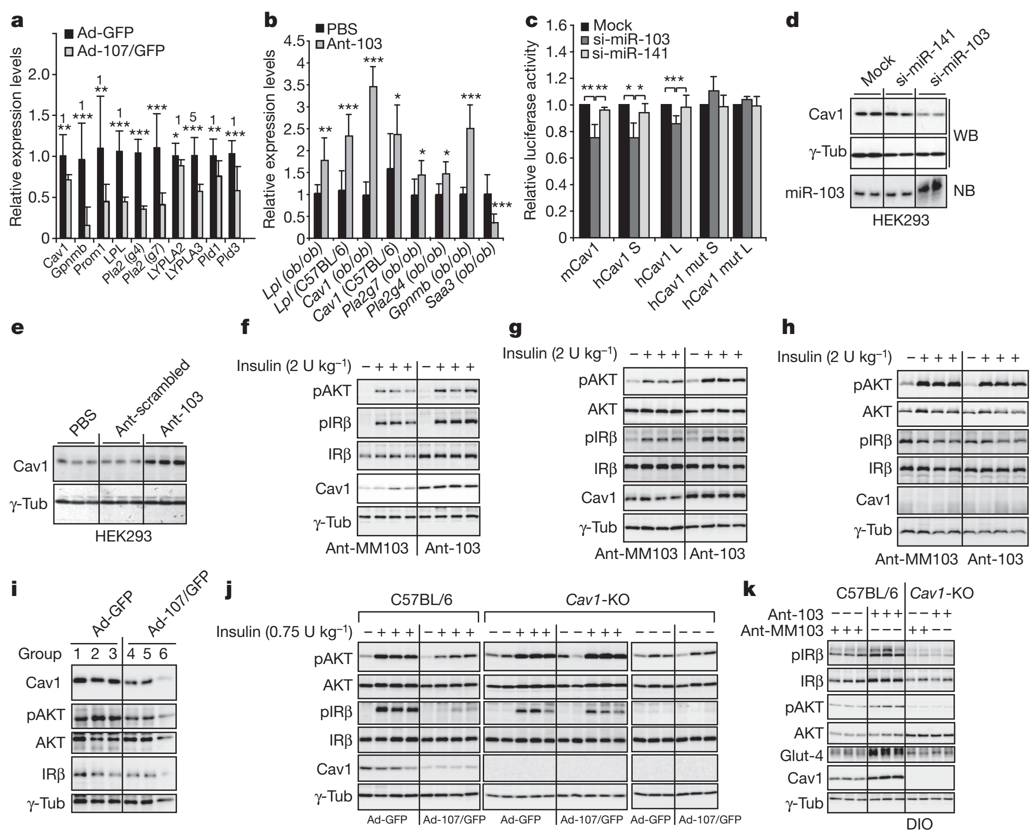


Figure 4 | Regulation of gene expression and insulin signalling by miR-103. **a**, **b**, Gene expression analysis (relative expression) in livers from C57BL/6J mice 10 d after injection with ad-GFP or ad-107/GFP (**a**, $n = 5$), or in fat from *ob/ob* mice injected with PBS or ant-103 (**b**, $n = 5$). **c**, Relative luciferase activity in HEK 293 cells transfected with reporter constructs containing the 3′ UTR of *Cav1*, co-transfected with si-miR-103, si-miR-141 (scrambled) or PBS (Mock). S, short; L, long; mut, mutant. **d**, Western blotting (WB) and northern blotting (NB) of HEK 293 cells transfected with PBS, si-miR-141 (scrambled) or si-miR-103. γ-Tub, γ-Tubulin. **e**, Immunoblotting of protein extracts from HEK 293 cells transfected with PBS, control ant-scrambled or ant-103. **f**–**h**, Immunoblotting of protein extracts from fat (**f**), liver (**g**) or muscle (**h**) of ant-MM103-injected or ant-103-injected *ob/ob* mice stimulated with 2 U kg^{-1} insulin for 8 min. **i**, Western blot analysis of perigonal fat pads in C57BL/6J mice surgically injected with ad-GFP or ad-107/GFP. Each lane represents a pool of two mice. **j**, Western blot analysis of liver extracts from C57BL/6J or *Cav1*-knockout (*Cav1*-KO) mice injected with ad-GFP or ad-107/GFP, stimulated with 0.75 U kg^{-1} insulin for 10 min after a 12 h fast. **k**, Immunoblotting of protein extracts from the fat of DIO C57BL/6J or DIO *Cav1*-KO mice 8 d after injection with ant-MM103 or ant-103 and stimulated for 8 min with 1.2 U kg^{-1} insulin after a 12 h fast. Animals were kept on a high-fat diet for 5 weeks before antagomir injection. Glut-4, glucose transporter-4. Means ± s.d. are shown for all panels. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

showed a reduction in Cav1 expression and decreased IRβ and pAkt1 levels (Fig. 4i). We also studied insulin signalling in the livers of mice with miR-107 overexpression. Cav1 and pAkt1 levels were diminished in the livers of wild-type mice infected with ad-miR-107, with no changes observed in IRβ protein levels (Fig. 4j). This result is in agreement with our findings showing that overexpression of miR-107 can induce hepatic insulin resistance, and with data from *Cav1*-null mice, which maintain normal IRβ levels in the liver but show reduced IRβ and pAkt1 levels in fat[19–21].

Finally, to investigate whether modulation of *Cav1* expression by miR-107 is important for the observed phenotypes, we overexpressed or silenced miR-103/107 in DIO *Cav1*-null mice. Whereas hepatic overexpression of miR-107 in wild-type mice led to impaired glucose tolerance, no significant effects on plasma glucose, glucose-tolerance, insulin-tolerance or pyruvate-tolerance tests were measured when *Cav1*-null mice were injected with ad-107/GFP and compared to ad-GFP-treated *Cav1*-null animals (Supplementary Fig. 9a–d). Furthermore, no molecular changes in insulin signalling events were detected in the two groups (Fig. 4j). Administration of ant-103 to DIO *Cav1*-null mice also did not affect glucose tolerance, insulin sensitivity or phosphorylation of insulin receptor and Akt1 upon insulin stimulation (Fig. 4k and Supplementary Fig. 9e–h). However, expression of lipolytic genes in the adipose tissue of *Cav1*-null mice was still responsive to treatment with ant-103, compared to ant-MM103 (Supplementary Fig. 9i), indicating that miR-103/107 also mediate some Cav1-independent metabolic effects.

Our findings show that miR-103 and miR-107 are negative regulators of insulin sensitivity. Their increased hepatic expression in rodents and humans with insulin resistance and hepatic steatosis indicates that they might contribute to the aetiology of diabetes. We also show that global miR-103/107 silencing causes increased insulin signalling in both liver and adipose tissue, although silencing of hepatic miR-103/107 expression in overt obese and insulin-resistant states is insufficient to reverse the metabolic abnormalities. This indicates that silencing of miR-103 in adipocytes is the dominant contributor to enhanced insulin sensitivity. One mechanism by which these miRNAs regulate insulin sensitivity is by targeting Cav1, thereby diminishing the number of insulin receptors in caveolae-enriched plasma membrane microdomains and reducing downstream insulin signaling. It is likely that Cav1 also mediates other effects that contribute to the phenotype because this protein has many functions in growth-factor signalling, endocytotic pathways and lipid regulation[22]. Our finding that silencing miR-103/107 in obese animals improves glucose homeostasis implicates these miRNAs as novel therapeutic targets for the treatment of diabetes.

METHODS SUMMARY

Animals. All mice were males and were maintained on a C57BL/6J background, on a 12-h light/dark cycle in a pathogen-free animal facility. Antagomirs at doses of 15 mg kg^{-1} in 0.2 ml total volume with PBS per injection were administered on two consecutive days through the tail vein of wild-type or *ob/ob* mice at between 6 and 8 weeks of age, or to 12-week-old DIO mice fed on a diet containing 60% fat (Pvolimi Kliba AG) for 8 weeks. Mice were injected with adenoviruses through the tail vein at 1×10^9 plaque-forming units in 0.2 ml PBS. Injection of ad-107/GFP and antagomirs did not effect food consumption compared to that of control-treated animals. Mice were killed 10 d after the adenovirus injection. All animal studies were approved by the Kantonale Veterinäramt Zürich.

Antagomirs. The single-stranded RNAs and modified RNA analogues used in this study consisted of 21–23 nucleotides with modifications as specified: antagomir-103, 5′-$u_s c_s$auagcccuguacaaugcu$_s g_s c_s u_s$-Chol-3′; antagomir-107, 5′-$u_s g_s$auagcccug uacaau gcu$_s g_s c_s u_s$-Chol-3′; MM-antagomir-103, 5′-$u_s g_s$acagccuuguaccaugcg$_s g_s c_s u_s$-Chol-3′; antagomir-124 (scrambled), 5′-$g_s g_s$cauucaccgcgugcc$_s u_s u_s a_s$-Chol-3′. The lower-case letters represent 2'-OMe-modified nucleotides; subscript 's' represents a phosphorothioate linkage; 'Chol' represents cholesterol linked through a hydroxyprolinol linkage[12].

Lipid nanoparticle formulations. Liver-targeting lipid nanoparticle formulations of antagomirs were prepared using the novel ionizable lipid DLin-KC2-DMA (ref. 22). Lipid nanoparticles were composed of DLin-KC2-DMA, distearoyl phosphatidylcholine, cholesterol and mPEG2000-DMG, used at the molar ratio 50:10:38.5:1.5. Antagomirs were formulated in lipid nanoparticles at a total lipid-to-antagomir weight ratio of approximately 11:1.

Generation of recombinant adenovirus. Ad-107/GFP was generated by inserting the PCR-amplified miRNA precursor sequence generated with primers 5′-AATACCCGCATGGAAGCAGGCTAA-3′ and 5′-AACATGTCTCAAGGA GAGGACGGT-3′ into a GFP-expressing shuttle vector, Ad5CMV K-NpA. Ad-GFP (ViraQuest) was used as a control.

Statistical analysis. Unless otherwise specified, all bars show mean ± s.d. Significance was calculated using student's *t*-test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 March 2010; accepted 13 April 2011.
Published online 8 June 2011.

1. Kahn, C. R. Knockout mice challenge our concepts of glucose homeostasis and the pathogenesis of diabetes. *Exp. Diabesity Res.* **4,** 169–182 (2003).
2. Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. Critical nodes in signalling pathways: insights into insulin action. *Nature Rev. Mol. Cell Biol.* **7,** 85–96 (2006).
3. Muoio, D. M. & Newgard, C. B. Mechanisms of disease: molecular and metabolic mechanisms of insulin resistance and β-cell failure in type 2 diabetes. *Nature Rev. Mol. Cell Biol.* **9,** 193–205 (2008).
4. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136,** 215–233 (2009).
5. Krützfeldt, J. & Stoffel, M. MicroRNAs: a new class of regulatory genes affecting metabolism. *Cell Metab.* **4,** 9–12 (2006).
6. Herrera, B. M. *et al.* Global microRNA expression profiles in insulin target tissues in a spontaneous rat model of type 2 diabetes. *Diabetologia* **53,** 1099–1109 (2010).
7. Anderson, N. & Borlak, J. Molecular mechanisms and therapeutic targets in steatosis and steatohepatitis. *Pharmacol. Rev.* **60,** 311–357 (2008).
8. Krützfeldt, J. *et al.* Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* **438,** 685–689 (2005).
9. Esau, C. *et al.* MicroRNA-143 regulates adipocyte differentiation. *J. Biol. Chem.* **279,** 52361–52365 (2004).
10. Kajimoto, K., Naraba, H. & Iwai, N. MicroRNA and 3T3–L1 pre-adipocyte differentiation. *RNA* **12,** 1626–1632 (2006).
11. Ortega, F. J. *et al.* MiRNA expression profile of human subcutaneous adipose and during adipocyte differentiation. *PLoS ONE* **5,** e9022 (2010).
12. Sun, T., Fu, M., Bookout, A. L., Kliewer, S. A. & Mangelsdorf, D. J. MicroRNA let-7 regulates 3T3–L1 adipogenesis. *Mol. Endocrinol.* **23,** 925–931 (2009).
13. Xie, H., Lim, B. & Lodish, H. F. MicroRNAs induced during adipogenesis that accelerate fat cell development are downregulated in obesity. *Diabetes* **58,** 1050–1057 (2009).
14. Goossens, G. H. The role of adipose tissue dysfunction in the pathogenesis of obesity-related insulin resistance. *Physiol. Behav.* **94,** 206–218 (2008).
15. Yamauchi, T. *et al.* The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity. *Nature Med.* **7,** 941–946 (2001).
16. Rothberg, K. G. *et al.* Caveolin, a protein component of caveolae membrane coats. *Cell* **68,** 673–682 (1992).
17. Nystrom, F. H., Chen, H., Cong, L. N., Li, Y. & Quon, M. J. Caveolin-1 interacts with the insulin receptor and can differentially modulate insulin signaling in transfected Cos-7 cells and rat adipose cells. *Mol. Endocrinol.* **13,** 2013–2024 (1999).
18. Yamamoto, M. *et al.* Caveolin is an activator of insulin receptor signaling. *J. Biol. Chem.* **273,** 26962–26968 (1998).
19. Otsu, K. *et al.* Caveolin gene transfer improves glucose metabolism in diabetic mice. *Am. J. Physiol. Cell Physiol.* **298,** C450–c456 (2009).
20. Cohen, A. W. *et al.* Caveolin-1-deficient mice show insulin resistance and defective insulin receptor protein expression in adipose tissue. *Am. J. Physiol. Cell Physiol.* **285,** C222–C235 (2003).
21. Cohen, A. W., Combs, T. P., Scherer, P. E. & Lisanti, M. P. Role of caveolin and caveolae in insulin signaling and diabetes. *Am. J. Physiol. Endocrinol. Metab.* **285,** E1151–E1160 (2003).
22. Parton, R. G. & Simons, K. The multiple faces of caveolae. *Nature Rev. Mol. Cell Biol.* **8,** 185–194 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank F. Preitner and B. Thorens for the hyperinsulinaemic euglycaemic clamp studies. M.T. was supported by a fellowship from the Juvenile Diabetes Research Foundation International. The work was supported in part by the Swiss National Science Foundation (SNF, LiverX), the European Community (SIROCCO, ERC and MetaboloMirs) and the Leducq Foundation.

Author Contributions M.T. and M.S. designed the experiments. M.T. performed the experiments and conducted the data analysis. J.H. and M.Z. performed the bioinformatic analysis. M.H.H. provided liver samples and participated in analysis of clinical data. B.B. synthesized antagomirs. A.A. provided liposomal formulations. M.T. and M.S. wrote the paper with input from all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.S. (stoffel@imsb.biol.ethz.ch).

METHODS

Animals. All mice were males and were maintained on a C57BL/6J background, on a 12-h light/dark cycle in a pathogen-free animal facility. Antagomirs at doses of 15 mg kg^{-1} in 0.2 ml total volume with PBS per injection were administered on two consecutive days through the tail vein of wild-type or *ob/ob* mice at between 6 and 8 weeks of age, or to 12-week-old DIO mice fed on a diet containing 60% fat (Pvolimi Kliba AG) for 8 weeks. Mice were injected with adenoviruses through the tail vein at 1×10^9 plaque-forming units in 0.2 ml PBS. Injection of ad-107/GFP and antagomirs did not effect food consumption compared to that of control-treated animals. Mice were killed 10 d after the adenovirus injection. All animal studies were approved by the Kantonale Veterinäramt Zürich.

BL/Adenovirus injection of fat. Ad-GFP or ad-107/GFP were injected into the perigonal fat at a concentration of 1×10^9 plaque-forming units in 40 µl PBS after surgical exposure. Animals were studied 8 d after injection.

Liver biopsies. Liver biopsy specimens from Caucasian patients were obtained during routine diagnostic work-up at the University Hospital, Basel. Blood samples were collected in the fasting state on the day of the liver biopsy for glucose and plasma-insulin measurements. Most study subjects did not take any medications. A specimen was frozen for research purposes if more than sufficient material was obtained for histopathological examination and if the patient gave his/her written informed consent in accordance with the Ethics Committee of Basel.

Antagomirs. The single-stranded RNAs and modified RNA analogues used in this study consisted of 21–23 nucleotides with modifications as specified: antagomir-103, 5′-$u_sc_s$auagcccuguacaaugc$u_sg_sc_su_s$-Chol-3′; antagomir-107, 5′-$u_sg_s$auagcccug uacaau gc$u_sg_sc_su_s$-Chol-3′; MM-antagomir-103, 5′-$u_sg_s$acagccuuguaccaugc$g_sg_s c_su_s$-Chol-3′; antagomir-124 (scrambled), 5′-$g_sg_s$cauucaccgcgugc$c_su_su_sa_s$-Chol-3′. The lower-case letters represent 2'-OMe-modified nucleotides; subscript 's' represents a phosphorothioate linkage; 'Chol' represents cholesterol linked through a hydroxyprolinol linkage[12].

Lipid nanoparticle formulations. Liver-targeting lipid nanoparticle formulations of antagomirs were prepared using the novel ionizable lipid DLin-KC2-DMA (ref. 22). Lipid nanoparticles were composed of DLin-KC2-DMA, distearoyl phosphatidylcholine, cholesterol and mPEG2000-DMG, used at the molar ratio 50:10:38.5:1.5. Antagomirs were formulated in lipid nanoparticles at a total lipid-to-antagomir weight ratio of approximately 11:1.

Generation of recombinant adenovirus. Ad-107/GFP was generated by inserting the PCR-amplified miRNA precursor sequence generated with primers 5′-AATACCCGCATGGAAGCAGGCTAA-3′ and 5′-AACATGTCTCAAGGA GAGGACGGT-3′ into a GFP-expressing shuttle vector, Ad5CMV K-NpA. Ad-GFP (ViraQuest), which does not contain a transgene, was used as a control.

RNA isolation and northern blotting analysis. 5–30 µg total RNA, isolated using Trizol reagent (Invitrogen), was separated at 15 W on 14% polyacrylamide gels containing formamide, as described in ref. 8.

Real-time PCR. Steady-state mRNA expression was measured by quantitative real-time PCR using the LightCycler 480 SYBR Green Master I Mix (Roche) with a Mx3005P Real-Time PCR System (Stratagene). Transcript levels were normalized to glyceraldehyde 3-phosphate dehydrogenase (GAPDH) or *36B4*, the gene encoding acidic ribosomal phosphoprotein P0 (RPLP0). Primer sequences for real-time PCRs are available on request. MiRNA levels were measured using TaqMan microRNA Assays (Applied Biosystems) and were normalized to U6 levels.

MicroRNA microarray. We used three diabetic groups (*ob/ob* or DIO) and three control groups of mice for each diabetic model. Prior to killing, elevated blood-glucose and insulin levels were confirmed in the diabetic mice. Total RNA, isolated and pooled from the livers of ten mice per group, was labelled using the miRCURY LNA microRNA Power Labelling Kit (Exiqon) and hybridized on miRNA arrays (miRXplore) that carry 1,194 DNA oligonucleotides with the reverse-complementary sequence of the mature RNAs. These arrays cover 728 human, 584 mouse, 426 rat and 122 viral miRNAs, each spotted on the arrays in quadruplicate. The Cy5-labelled miRNAs were normalized to a reference pool of miRNAs that were simultaneously labelled with Cy3. All the data are represented as ratios of logarithmic values between the diabetic and healthy animals ± s.d.

Assay of luciferase activity and cell culture transfection. 3′ UTR sequences were PCR-amplified with specific primers, followed by attB adaptor PCR. Sequences were cloned into the pDONR221 entry vector using BP Clonase (Invitrogen) and then cloned behind the stop codon of firefly luciferase in the dual renilla/firefly luciferase pEM393 destination vector (gift from E. Miska). HEK 293 cells cultured in 24-well plates were transfected in quadruplicate using Lipofectamine 2000 (Invitrogen) with 100 ng of the final construct per well, together with PBS or 50 nmol of either control or si-103 double-stranded siRNA (Sigma). Cells were collected 42–48 h after transfection and assayed using the Dual-Luciferase Reporter Assay System (Promega). Results were normalized to the renilla luciferase control and expressed relative to the average value of the control, which was treated with PBS. HEK 293 cells were transfected with antagomirs at a concentration of 5.5 µg ml^{-1} of medium.

Computer tomography. Animals were scanned using an animal CT-Scanner (LaTheta) at 1 mm intervals from the head to the base of the tail. Images were analysed using the LaTheta Software.

Isolation of stromal-vascular fraction and primary adipocytes. Primary adipocytes and the stromal-vascular fraction from subcutaneous and visceral fat were prepared as previously described[23,24]. Adipocyte differentiation was induced with insulin, dexamethasone, isobutylmethylxanthine and rosiglitazone when stromal-vascular cells were 80% confluent[24]. Cells were treated with antagomirs at a concentration of 5.5 µg ml^{-1} during the induction period on days 2 and 3.

Automated analysis of adipocyte differentiation. Differentiated cells were fixed with 5% formaldehyde before staining with boron-dipyrromethene (BODIPY) for lipid droplets, Hoechst for nuclei and Syto60 for cytosolic staining (Invitrogen). A total of 25 pictures per well were taken with an automated microscope imaging system (CellWorx). Pictures were analysed using Cell Profiler Software.

Glucose uptake. [^{14}C]-Spiked glucose uptake, with or without 20 nM insulin stimulation, was measured as previously described[24].

Hyperinsulinaemic-euglycaemic clamp studies. Clamps were performed on 8-week-old *ob/ob* mice weighing 40 g. An indwelling catheter for infusion of insulin and glucose was placed into the left femoral vein under isoflurane anaesthesia. Mice were allowed to recover for 6–8 d, until they had regained 95–100% of their initial body weight. After a 5 h fast, a 180 min hyperinsulinaemic-euglycaemic clamp study was conducted in awake, freely moving mice, as previously described[25,26].

Adipocyte size. Haematoxylin and eosin staining of 10 µm slices of adipose tissue fixed in 5% paraformaldehyde was performed according to standard procedures[27] and images were analysed using Cell Profiler Software. At least 2,000 adipocytes were measured per animal to determine adipocyte size.

Glucose-, insulin- and pyruvate-tolerance tests. Glucose-, insulin- and pyruvate-tolerance tests were performed by intraperitoneal injection of glucose (2 g kg^{-1}), insulin (0.75 units kg^{-1}, 1 unit kg^{-1} or 2 units kg^{-1}, as indicated in the figures) or pyruvate (2 g kg^{-1}) after an overnight fast for glucose and pyruvate or a 6 h fast for insulin. Blood glucose levels were measured before injection (time 0) and at 15, 30, 60 and 120 min after injection.

Antibodies. The antibodies used were mouse monoclonal anti-γ-tubulin (Sigma-Aldrich), rabbit polyclonals anti-IRβ (C-19):sc-711; anti-pIRβ (Tyr1162/1163):sc25103, anti-caveolin-1 (N20):sc-894 (Santa Cruz Biotechnology), anti-pAKT and anti-AKT (Cell Signaling).

Statistical analysis. All bars show mean ± s.d. Significance was calculated using student's *t*-test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

23. Semple, S. C. *et al.* Rational design of cationic lipids for siRNA delivery. *Nature Biotechnol.* **28,** 172–176 (2010).
24. Hansen, L. H., Madsen, B., Teisner, B., Nielsen, J. H. & Billestrup, N. Characterization of the inhibitory effect of growth hormone on primary preadipocyte differentiation. *Mol. Endocrinol.* **12,** 1140–1149 (1998).
25. Tozzo, E., Shepherd, P. R., Gnudi, L. & Kahn, B. B. Transgenic GLUT-4 overexpression in fat enhances glucose metabolism: preferential effect on fatty acid synthesis. *Am. J. Physiol.* **268,** E956–E964 (1995).
26. Minehira, K. *et al.* Blocking VLDL secretion causes hepatic steatosis but does not affect peripheral lipid stores or insulin sensitivity in mice. *J. Lipid Res.* **49,** 2038–2044 (2008).
27. Preitner, F., Mody, N., Graham, T. E., Peroni, O. D. & Kahn, B. B. Long-term Fenretinide treatment prevents high-fat diet-induced obesity, insulin resistance, and hepatic steatosis. *Am. J. Physiol. Endocrinol. Metab.* **297,** E1420–E1429 (2009).

doi:10.1038/nature10113

Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia

Xose S. Puente[1], Magda Pinyol[2], Víctor Quesada[1], Laura Conde[3], Gonzalo R. Ordóñez[1], Neus Villamor[3], Georgia Escaramis[4], Pedro Jares[3], Sílvia Beà[3], Marcos González-Díaz[5], Laia Bassaganyas[4], Tycho Baumann[6], Manel Juan[7], Mónica López-Guerra[3], Dolors Colomer[3], José M. C. Tubío[4,8], Cristina López[3], Alba Navarro[3], Cristian Tornador[4], Marta Aymerich[3], María Rozman[3], Jesús M. Hernández[5], Diana A. Puente[1], José M. P. Freije[1], Gloria Velasco[1], Ana Gutiérrez-Fernández[1], Dolors Costa[3], Anna Carrió[3], Sara Guijarro[3], Anna Enjuanes[3], Lluís Hernández[3], Jordi Yagüe[7], Pilar Nicolás[9], Carlos M. Romeo-Casabona[9], Heinz Himmelbauer[10], Ester Castillo[10], Juliane C. Dohm[10], Silvia de Sanjosé[11], Miguel A. Piris[12], Enrique de Alava[5], Jesús San Miguel[5], Romina Royo[13], Josep L. Gelpí[13], David Torrents[13], Modesto Orozco[13], David G. Pisano[14], Alfonso Valencia[14], Roderic Guigó[15], Mónica Bayés[16], Simon Heath[16], Marta Gut[16], Peter Klatt[17], John Marshall[18], Keiran Raine[18], Lucy A. Stebbings[18], P. Andrew Futreal[18], Michael R. Stratton[18], Peter J. Campbell[18], Ivo Gut[16], Armando López-Guillermo[6], Xavier Estivill[4], Emili Montserrat[6], Carlos López-Otín[1]* & Elías Campo[3]*

Chronic lymphocytic leukaemia (CLL), the most frequent leukaemia in adults in Western countries, is a heterogeneous disease with variable clinical presentation and evolution[1,2]. Two major molecular subtypes can be distinguished, characterized respectively by a high or low number of somatic hypermutations in the variable region of immunoglobulin genes[3,4]. The molecular changes leading to the pathogenesis of the disease are still poorly understood. Here we performed whole-genome sequencing of four cases of CLL and identified 46 somatic mutations that potentially affect gene function. Further analysis of these mutations in 363 patients with CLL identified four genes that are recurrently mutated: notch 1 (*NOTCH1*), exportin 1 (*XPO1*), myeloid differentiation primary response gene 88 (*MYD88*) and kelch-like 6 (*KLHL6*). Mutations in *MYD88* and *KLHL6* are predominant in cases of CLL with mutated immunoglobulin genes, whereas *NOTCH1* and *XPO1* mutations are mainly detected in patients with unmutated immunoglobulins. The patterns of somatic mutation, supported by functional and clinical analyses, strongly indicate that the recurrent *NOTCH1*, *MYD88* and *XPO1* mutations are oncogenic changes that contribute to the clinical evolution of the disease. To our knowledge, this is the first comprehensive analysis of CLL combining whole-genome sequencing with clinical characteristics and clinical outcomes. It highlights the usefulness of this approach for the identification of clinically relevant mutations in cancer.

To gain insights into the molecular alterations that cause CLL, we performed whole-genome sequencing of four cases representative of different forms of the disease: two cases, CLL1 and CLL2, with no mutations in the immunoglobulin genes (*IGHV*-unmutated) and two cases, CLL3 and CLL4, with mutations in these genes (*IGHV*-mutated) (Supplementary Table 1 and Supplementary Information). We used a combination of whole-genome sequencing and exome sequencing, as well as long-insert paired-end libraries, to detect variants in chromosomal structure (Supplementary Fig. 1 and Supplementary Tables 2–5). We obtained more than 99.7% concordance between whole-genome sequencing calls and genotyping data, indicating that the coverage and parameters used were sufficient to detect most of the sequence variants in these samples (Supplementary Information). We detected about 1,000 somatic mutations per tumour in non-repetitive regions (Fig. 1a, Supplementary Fig. 2 and Supplementary Table 6). These numbers of somatic mutations were lower than the numbers in melanoma and lung carcinoma[5,6], but in agreement with previous estimates of less than one mutation per megabase (Mb) for leukaemias[7]. The most common substitution was the transition G>A/C>T, usually occurring in a CpG context (Fig. 1b and Supplementary Fig. 2). We also detected marked differences in the mutation pattern between CLL samples and these differences were associated with tumour subtype (Fig. 1b). Thus, *IGHV*-mutated cases showed a higher proportion of A>C/T>G mutations than cases with unmutated *IGHV* ($16 \pm 0.2\%$ versus $6.2 \pm 0.1\%$). The base preceding the adenine in A to C transversions showed an overrepresentation of thymine, when compared to the prevalence expected from its representation in non-repetitive sequences in the wild-type genome ($P < 0.001$, Fig. 1c), and there were fewer A to C substitutions at GpA dinucleotides than would be expected by chance ($P < 0.001$). These differences between CLL subtypes might reflect the molecular mechanisms implicated in their respective development. The pattern and context of mutations are consistent with their being introduced by the error-prone polymerase η during somatic hypermutation in immunoglobulin genes[8]. This indicates that polymerase η could contribute to the high frequency of A > T to C > G transversions in cases with *IGHV*-mutated. It also extends the differences observed between these two CLL subtypes to the genomic level.

We classified the somatic mutations into three different classes according to their potential functional effect (Supplementary Information). We also searched for small insertions and deletions (indels) in coding regions: we found and validated five somatic indels, which caused frameshifts in protein-coding regions (Supplementary Table 7).

[1]Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología, Universidad de Oviedo, 33006 Oviedo, Spain. [2]Unidad de Genómica, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Spain. [3]Unidad de Hematopatología, Servicio de Anatomía Patológica, Hospital Clínic, Universitat de Barcelona, IDIBAPS, 08036 Barcelona, Spain. [4]Genes and Disease Programme, Center for Genomic Regulation, Pompeu Fabra University (CRG-UPF), 08003 Barcelona, Spain. [5]Servicio de Hematología, Hospital Universitario, Centro de Investigación del Cáncer-IBMCC (USAL-CSIC), Universidad de Salamanca, 37007 Salamanca, Spain. [6]Servicio de Hematología, Hospital Clínic, IDIBAPS, Universidad de Barcelona, 08036 Barcelona, Spain. [7]Servicio de Inmunología, Hospital Clínic, IDIBAPS, 08036 Barcelona, Spain. [8]Hospital Clínico Universitario de Santiago de Compostela, 15706 Santiago de Compostela, Spain. [9]Cátedra Inter-Universitaria de Derecho y Genoma Humano, Universidad de Deusto, Universidad del País Vasco, 48007 Bilbao, Spain. [10]Ultrasequencing Unit, Center for Genomic Regulation, Pompeu Fabra University (CRG-UPF), 08003 Barcelona, Spain. [11]Unidad de Infecciones y Cáncer, Institut Català de Oncologia-IDIBELL, 08907 Hospitalet de Llobregat, Spain. [12]Molecular Pathology Programme, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain. [13]Programa Conjunto de Biología Computacional, Barcelona Supercomputing Center (BSC), Institut de Recerca Biomèdica (IRB), Spanish National Bioinformatics Institute, Universitat de Barcelona, 08028 Barcelona, Spain. [14]Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Spanish National Bioinformatics Institute, 28029 Madrid, Spain. [15]Genomic Bioinformatics Programme, Center for Genomic Regulation, Pompeu Fabra University (CRG-UPF), Spanish National Bioinformatics Institute, 08003 Barcelona, Spain. [16]Centro Nacional de Análisis Genómico, Parc Científic de Barcelona, 08028 Barcelona, Spain. [17]Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain. [18]Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK.
*These authors contributed equally to this work.

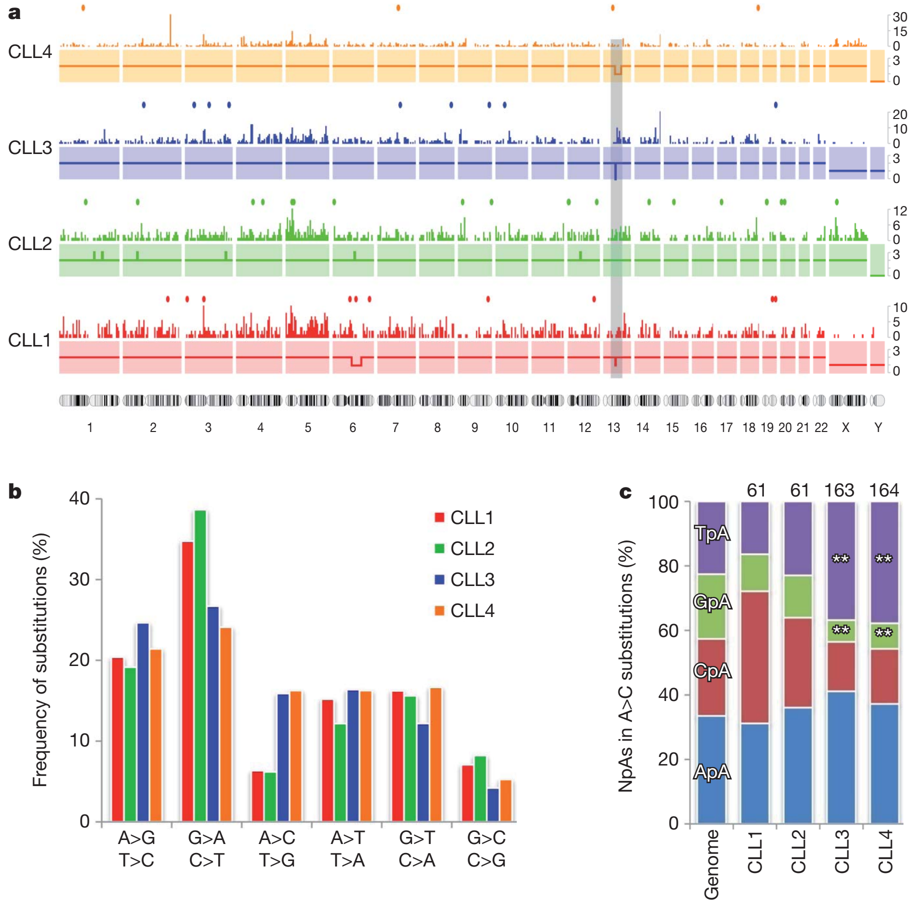


Figure 1 | Profile of somatic mutations in four CLL genomes. a, Distribution of somatic alterations. For each tumour genome, copy number (solid lines), density of mutations per 5-Mb window (bars) and protein-coding mutations (dots) are shown. The shaded rectangle indicates the location of the 13q14 deletion that was present in three of the four CLL cases. Chromosome numbers are listed below the four profiles. **b**, Frequency of substitutions in each CLL tumour for the six possible classes of mutation. **c**, Distribution of the four possible NpA dinucleotides for the A to C transversion in each tumour genome, compared with the expected distribution across the genome. The total number of A to C substitutions per case is indicated at the top (**, $P < 0.001$).

We identified 46 mutations that changed the protein-coding sequences of 45 genes in the four patients analysed (Supplementary Table 7). None of these nucleotide substitutions had been previously linked to CLL and among the five indel mutations, only one, in *NOTCH1* (p.P2515Rfs*4), had been previously found in various lymphoid malignancies, including CLL[9,10]. To determine whether any of these 45 genes was mutated in more than one CLL case, we analysed an initial validation set of 169 CLL patients. We focused on the 26 genes that are expressed at the RNA level in CLL cells (Supplementary Table 7) because mutations in expressed genes are more likely to have a biological effect than those in non-expressed genes. We used a pooled-sequencing strategy that led us to identify four genes with at least one additional mutation in the validation series: these were *NOTCH1*, *MYD88*, *XPO1* and *KLHL6* (Table 1 and Supplementary Information).

Analysis of additional CLL cases revealed that the deletion of a CT dinucleotide in *NOTCH1* (p.P2515Rfs*4) was found in 29 of 255 patients and two additional mutations in the same region were also found (p.Q2503* and p.F2482Ffs*2) (Fig. 2a, b). Accordingly, *NOTCH1* is mutated in 12% of CLL patients (Supplementary Table 8). These mutations generate a premature stop codon, resulting in a NOTCH1 protein lacking the C-terminal domain, which contains a PEST sequence (a sequence rich in proline, glutamic acid, serine and threonine) (Fig. 2a). Removal of this region results in the accumulation of an active protein isoform in the mutated CLL cells (Fig. 2c and Supplementary Fig. 3). NOTCH1 is constitutively expressed in CLL[11], but the *NOTCH1* mutations identified herein generate a more stable and active isoform of the protein. Gene expression analysis of ten *NOTCH1*-mutated and 49 unmutated CLL cases revealed a high

Table 1 | Genes recurrently mutated in chronic lymphocytic leukaemia

Gene	Protein	Mutation	Mutated cases / total	Overall frequency (%)	Frequency in *IGHV*-unmutated (%)	Frequency in *IGHV*- mutated (%)
NOTCH1	Notch 1	P2515Rfs*4	29/255	12.2	20.4	7
		Q2503*	1/255			
		F2482Ffs*2	1/255			
MYD88	Myeloid differentiation primary response gene 88	L265P	9/310	2.9	0.8	5.6
XPO1	Exportin 1	E571K	3/165	2.4	4.6	0
		E571G	1/165			
KLHL6	Kelch-like 6	F49L/L65P	3/160	1.8	0	4.5
		L90F				
		L58P/T64A/Q81P				

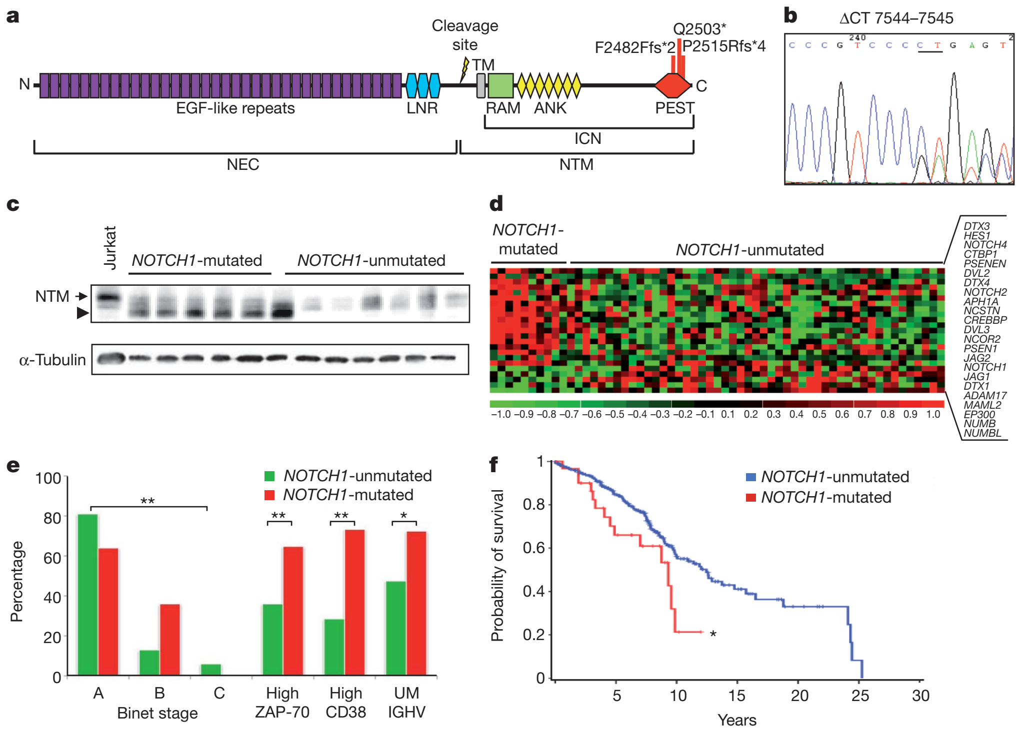


Figure 2 | Mutational and functional analysis of *NOTCH1* in CLL. **a**, Schematic representation of human NOTCH1, showing the main domains and locations of the three different somatic mutations identified in CLL. NEC, NOTCH1 extracellular subunit; NTM, NOTCH1 transmembrane subunit; ICN, intracellular domain of NOTCH1; LNR, Lin-12 NOTCH repeats; RAM, RAM domain; ANK, ankyrin repeat domain; PEST, PEST domain. **b**, Electropherogram showing the heterozygous CT deletion recurrently identified in CLL. **c**, Western blot showing NOTCH1 protein levels in CLL cases with or without the NOTCH1 p.P2515Rfs*4 mutation, and in Jurkat cells as a control. The arrow indicates the band corresponding to the NTM; the large arrowhead indicates the smaller band corresponding to the mutant form. **d**, Heat map showing the 23 genes of the NOTCH1 pathway that are differentially expressed in *NOTCH1*-mutated versus non-mutated CLL. **e**, Distribution of disease stage (Binet), ZAP-70 expression status, CD38 expression status and *IGHV* mutational status (UM, unmutated *IGHV*) in patients with or without mutations in *NOTCH1* (*, $P < 0.02$; **, $P < 0.01$). **f**, Actuarial probability of overall survival of CLL patients with mutated or unmutated *NOTCH1* (*, $P = 0.03$).

number of differentially expressed genes ($n = 542$, false discovery rate <0.05; Supplementary Table 9). Likewise, in a gene-set analysis, we found that there was significant differential expression of the NOTCH1 signalling pathway[12] and two metabolic pathways (oxidative phosphorylation and glycolysis/gluconeogenesis). This is consistent with the NOTCH1-mediated activation of multiple biosynthetic routes in T acute lymphoblastic leukaemia[13]. When the differential expression of individual genes from the NOTCH1 pathway was analysed, 23 of the 46 genes assigned to this pathway[12] showed a significant differential expression ($P < 0.05$) in *NOTCH1*-mutated CLL (Fig. 2d). *NOTCH1*-mutated patients had a more advanced clinical stage at diagnosis, more adverse biological features and an overall survival that was significantly shorter than those with *NOTCH1* unmutated (10-yr overall survival: 21% versus 56%, $P = 0.03$; Fig. 2e, f). *NOTCH1*-mutated CLL also underwent transformation into diffuse large B-cell lymphoma more frequently than *NOTCH1*-unmutated CLL (7 of 31 cases, 23%, versus 3 of 224 cases, 1.3%; $P < 0.001$). The same *IGHV* clonal rearrangement and *NOTCH1* mutation were found in the CLL and corresponding transformed diffuse large B-cell lymphoma of the four cases studied, indicating a clonal relationship of both components.

A recurrent mutation (p.L265P) in the *MYD88* gene (Fig. 3a, b) was also identified in 9 of 310 CLL patients (2.9%). During revision of this manuscript, the same mutation has been identified in different lymphomas[14], highlighting its relevance in the pathogenesis of lymphoid neoplasias. This protein participates in the signalling pathways of interleukin-1 and Toll-like receptors during the immune response[15]. MyD88 immunoprecipitation from CLL cells with the p.L265P mutation resulted in the co-immunoprecipitation of large amounts of IRAK1, in contrast to cells lacking this mutation (Fig. 3c). Other effectors of this signalling pathway, including STAT3, IκBα and NF-κB p65 subunit, showed higher phosphorylation in *MYD88*-mutated than in unmutated CLL cells (Fig. 3d, e) and there was an increased DNA-binding activity of NF-κB in *MYD88*-mutated cells (Supplementary Fig. 4). These data support the hypothesis that the *MYD88* p.L265P mutation constitutes an activating mutation of this novel proto-oncogene[14,16]. Stimulation of interleukin-1 receptor or Toll-like receptors in *MYD88*-mutated CLL cells induced the secretion of 5-fold to 150-fold higher levels of interleukin 1 receptor antagonist (IL1RN, also known as IL1RA), interleukin 6 and chemokine (C-C motif) ligands 2, 3 and 4 (CCL2, CCL3 and CCL4), when compared to the secretion of these cytokines by *MYD88*-unmutated CLLs. Cytokine secretion was elevated in *MYD88*-mutated cells in response to stimulation of at least four of the eight TLRs tested. No response was observed in lymphocytes carrying the inactivating *MYD88* mutation E52DEL (Fig. 3f and Supplementary Fig. 5). The high production of these cytokines has been implicated in the recruitment of macrophages and T lymphocytes by CLL cells, creating a favourable niche for their survival[17]. Moreover, activation of Toll-like receptors in CLL cells promotes the proliferation of tumour cells and protects them from spontaneous apoptosis[18]. Patients with *MYD88*-mutated CLL were diagnosed at a younger age than those with wild-type *MYD88* (median 43 yr, range 38–63, versus median 63 yr, range 27–94; $P < 0.001$) and the disease presented with a more advanced clinical stage (Fig. 3g), although no differences were observed in progression or survival rates. Notably, almost all patients with the *MYD88* p.L265P mutation (seven of the eight evaluated) belonged to the *IGHV*-mutated group.

We also identified four cases with mutations in the same codon of the exportin 1 gene (*XPO1*; p.E571K and p.E571G). Exportin 1 is implicated in the nuclear export of proteins and mRNAs in yeast, including members of the MAP kinase pathway[19]. The fact that the

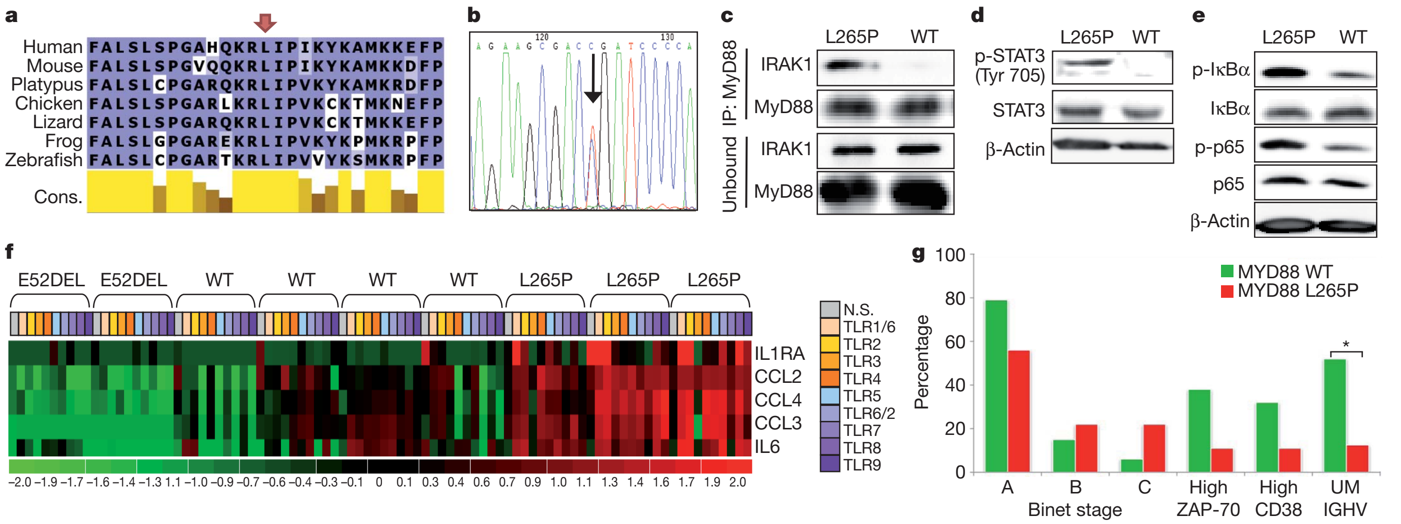


Figure 3 | Mutational and functional analysis of *MYD88* in CLL. **a**, Multiple sequence alignment of MyD88 around the mutated residue (arrow) in different species. Cons., degree of conservation. **b**, Electropherogram showing the recurrent heterozygous p.L265P *MYD88* mutation (arrow) detected in CLL. **c**, Cell extracts from a *MYD88*-mutated CLL (L265P) and a *MYD88*-unmutated CLL (WT) were immunoprecipitated with anti-MyD88 antibody. The immunoprecipitated and unbound fractions were analysed by western blot using anti-IRAK1 and anti-MyD88 antibodies. **d**, Western blot analysis of phosphorylated STAT3 (p-STAT3 (Tyr 705)) and total STAT3 in cell extracts from *MYD88*-mutated or unmutated CLL tumour cells. β-Actin was used as a control to show equal loading. **e**, Western blots showing phosphorylated IκBα (p-IκBα), total IκBα, phosphorylated p65 subunit of NF-κB (p-p65) and total p65 subunit of NF-κB in cell extracts from *MYD88*-mutated or unmutated CLL tumour cells. **f**, Heat map representing the cytokine levels secreted by B cells from eight individuals after Toll-like receptor stimulation. Only the five cytokines that showed the most significant differences between *MYD88*-mutated and *MYD88*-unmutated CLL are shown. 'E52DEL' indicates B cells from two patients with an inactivating *MYD88* mutation, 'WT' corresponds to tumour cells from CLL patients without *MYD88* mutation and 'L265P' indicates tumour cells from patients carrying a mutated *MYD88*. The stimulation experiments for each of the Toll-like receptors (TLRs) are represented in different colours. NS, no stimulus. **g**, Distribution of disease stage (Binet), ZAP-70 expression status, CD38 expression status and *IGHV* mutational status (UM, unmutated *IGHV*) in patients according to the presence or absence of p.L265P *MYD88* mutation (*, $P < 0.03$).

same residue is mutated in four CLL cases and is part of a highly conserved region (Supplementary Fig. 6) indicates that the mutation affects XPO1 activity. Notably, all four cases with mutations in *XPO1* belonged to the *IGHV*-unmutated subtype and two of them also had the p.P2515Rfs*4 mutation in *NOTCH1*, indicating that both mutations could have synergic effects in CLL development.

We identified three patients carrying a total of six mutations (F49L/L65P, L90F and L58P/T64A/Q81P) in the gene encoding kelch-like protein 6 (*KLHL6*), which is implicated in the formation of the germinal centre during B cell maturation[20]. All six mutations were clustered between residues 49 and 90 (Supplementary Fig. 7). The presence of several point mutations in *cis*, located near the transcriptional start site of a gene that is highly expressed in the germinal centre, is a characteristic feature of somatic hypermutation. In fact, all three patients had CLL with mutated *IGHV*. Although somatic hypermutation occurs mainly in *IGHV* regions, other proto-oncogenes, including *BCL6*, *MYC* and *PIM1*, are mutated by somatic hypermutation in different lymphomas[21]. However, only *BCL6* has been previously shown to be hypermutated by this mechanism in CLL[21]. Our data show that *KLHL6* is probably also a target of somatic hypermutation in *IGHV*-mutated patients, although its precise contribution to the oncogenic process in CLL remains to be determined.

In addition to these four genes, we identified a series of large genomic alterations that were previously reported[2]. They included the deletion, in three cases, of the 13q14 region[22], and a 40-Mb deletion in chromosome 6q14–q22 (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 5). Finally, in one patient we detected a p.P281R mutation in the cyclin D2 gene (*CCND2*), which resulted in the accumulation of cyclin D2 in tumour cells (Supplementary Fig. 8). This finding, together with the high conservation of this residue and the identification of mutations in the equivalent residue of cyclin D1 (*CCND1*) in endometrial cancer[23], indicates that this *CCND2* mutation could be a driver contributing to the development of CLL in this patient. The finding illustrates the putative relevance of non-recurrent mutations for the pathogenesis of CLL.

The International Cancer Genome Consortium project was founded on the concept that sequencing of cancer genomes could reshape our understanding of cancer biology, with direct implications for clinical translation[24]. Our study of four CLL genomes underscores this transformative potential, although additional studies will be necessary to translate these findings to the clinic. We have identified four recurrently mutated genes and provided novel insights into the mechanisms by which leukaemic cells recruit, instruct and coordinate a tumour microenvironment. Currently, the biological identification of different subgroups of CLL is based on markers such as *IGHV* mutational status, cytogenetics, ZAP-70 expression or CD38 expression, which are not fundamental agents in the leukaemic process. The classification of patients based on genomic drivers of the disease is conceptually appealing, as shown by our demonstration that *NOTCH1* and *MYD88* mutations identify distinct subgroups of patients with particular clinical and biological features. Furthermore, we provide functional evidence that both *NOTCH1* and *MYD88* mutations are activating events and potential therapeutic targets. The potential to personalize therapeutic choices for patients on the basis of the genomic architecture of their cancers is the long-term aspiration for studies such as this, combining whole-genome sequencing, functional studies and clinical analysis of patients with cancer.

METHODS SUMMARY

Four patients with CLL, who had given informed consent for sample collection and analysis, were studied. Tumour samples were obtained before treatment and tumour cells were separated from non-tumour cells by immunomagnetic depletion of T cells, natural killer cells, monocytes and granulocytes (Supplementary Information). Tumour cell purity was ≥98% as assessed by flow cytometry. Normal blood cells from the same patient were obtained after treatment, resulting in no detectable, or less than 0.05%, tumour cell contamination, as assessed by flow cytometry. Additional samples from 363 patients were obtained for clinical validation. Protocols for long-insert and short-insert library construction and for massively parallel paired-end sequencing have been described elsewhere (ref. 25 and Supplementary Information). Genotyping and copy number analysis were performed using the Affymetrix SNP6.0, Agilent 1M and Illumina OmniQuad arrays

on the same cases used for whole-genome sequencing. For the validation of candidate genes in a set of 169 additional CLL patients, we used a combination of PCR amplification and Illumina sequencing in pooled samples, resulting in efficient identification of germline and somatic mutations (Supplementary Information). Sequencing data were aligned to the human reference genome (GRCh37) using Burrows–Wheeler alignment (BWA)[26] and somatic substitutions were identified using Sidrón, a probabilistic binomial model that uses genotyping data to calibrate sequencing error per sample. Functional analyses of the identified mutations were performed using cryopreserved primary tumour cells. For gene expression analysis, RNA was purified from tumour cells and analysed using the HU133 plus 2.0 GeneChip (Affymetrix). For immunoprecipitation and western blotting, CLL cell extracts were prepared and detected using the indicated antibodies (Supplementary Information). For Toll-like receptor stimulation of CLL cells, the Human TLR1–9 agonist kit (InvivoGen) was used.

Received 16 November 2010; accepted 6 April 2011.
Published online 5 June 2011.

1. Rozman, C. & Montserrat, E. Chronic lymphocytic leukemia. *N. Engl. J. Med.* **333,** 1052–1057 (1995).
2. Zenz, T., Mertens, D., Kuppers, R., Dohner, H. & Stilgenbauer, S. From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nature Rev. Cancer* **10,** 37–50 (2010).
3. Damle, R. N. *et al.* Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* **94,** 1840–1847 (1999).
4. Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G. & Stevenson, F. K. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **94,** 1848–1854 (1999).
5. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463,** 191–196 (2010).
6. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463,** 184–190 (2010).
7. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153–158 (2007).
8. Spencer, J. & Dunn-Walters, D. K. Hypermutation at A-T base pairs: the A nucleotide replacement spectrum is affected by adjacent nucleotides and there is no reverse complementarity of sequences flanking mutated A and T nucleotides. *J. Immunol.* **175,** 5170–5177 (2005).
9. Weng, A. P. *et al.* Activating mutations of *NOTCH1* in human T cell acute lymphoblastic leukemia. *Science* **306,** 269–271 (2004).
10. Sportoletti, P. *et al.* *NOTCH1* PEST domain mutation is an adverse prognostic factor in B-CLL. *Br. J. Haematol.* **151,** 404–406 (2010).
11. Rosati, E. *et al.* Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. *Blood* **113,** 856–865 (2009).
12. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38,** D355–D360 (2010).
13. Palomero, T. *et al.* NOTCH1 directly regulates *c-MYC* and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc. Natl Acad. Sci. USA* **103,** 18261–18266 (2006).
14. Ngo, V. N. *et al.* Oncogenically active *MYD88* mutations in human lymphoma. *Nature* **470,** 115–119 (2011).
15. O'Neill, L. A. & Bowie, A. G. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nature Rev. Immunol.* **7,** 353–364 (2007).
16. Coste, I. *et al.* Dual function of MyD88 in RAS signaling and inflammation, leading to mouse and human cell transformation. *J. Clin. Invest.* **120,** 3663–3667 (2010).
17. Burger, J. A. *et al.* High-level expression of the T-cell chemokines CCL3 and CCL4 by chronic lymphocytic leukemia B cells in nurselike cell cocultures and after BCR stimulation. *Blood* **113,** 3050–3058 (2009).
18. Muzio, M. *et al.* Expression and function of toll like receptors in chronic lymphocytic leukaemia cells. *Br. J. Haematol.* **144,** 507–516 (2009).
19. Ferrigno, P., Posas, F., Koepp, D., Saito, H. & Silver, P. A. Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin β homologs NMD5 and XPO1. *EMBO J.* **17,** 5606–5614 (1998).
20. Kroll, J. *et al.* The BTB-kelch protein KLHL6 is involved in B-lymphocyte antigen receptor signaling and germinal center formation. *Mol. Cell. Biol.* **25,** 8531–8540 (2005).
21. Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412,** 341–346 (2001).
22. Aqeilan, R. I., Calin, G. A. & Croce, C. M. miR-15a and miR-16–1 in cancer: discovery, function and future perspectives. *Cell Death Differ.* **17,** 215–220 (2010).
23. Moreno-Bueno, G. *et al.* Cyclin D1 gene (*CCND1*) mutations in endometrial cancer. *Oncogene* **22,** 6115–6118 (2003).
24. Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464,** 993–998 (2010).
25. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).
26. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by the Spanish Ministry of Science and Innovation (MICINN) through the Instituto de Salud Carlos III (ISCIII) and Red Temática de Investigación del Cáncer (RTICC) del ISCIII. C.L.-O. is an Investigator of the Botin Foundation and D.T., of the ICREA program. We thank E. Santos for his support of this project, A. Carracedo and J. Benítez for genotyping studies, C. Fortuny for the supply of samples and N. Villahoz and M. C. Muro for their work in the coordination of the CLL-ICGC Consortium. We are also grateful to all patients with CLL who participated in this study.

Author Contributions X.S.P., V.Q., G.R.O., J.M.P.F., G.V., and A.G.-F. developed the bioinformatic algorithms for analysis of sequence data, implemented SMIPS for clinical validation and performed functional studies. M.P., L.C., P.J., M.J., M.L.-G., D.Colomer, S.G., C.L., A.N. and J.Y. were responsible for downstream validation analysis and functional studies. S.B., A.C., G.E., L.B., J.M.C.T., J.C.D. and C.T. studied structural variants. M.G., M.B., S.H., D.A.P., H.H., E.C., J.M., K.R. and L.A.S. were responsible for generating libraries, performing exome capture and running sequencers. P.N., C.M.R-C. and M.A. prepared and supervised the bioethical requirements. S.S. designed the epidemiologic study. P.K. gave conceptual advice and revised the manuscript. N.V., T.B., D.Costa, A.C., A.E., L.H., M.A.P., E.A., J.S.M., E.M., A.L.-G. and E.C. performed clinical and biological studies. M.R., N.V., J.M.H. and M.G.-D. were the pathologists who reviewed and confirmed the diagnoses. R.R., J.L.G., M.O., D.T., D.G.P., A.V. and R.G. were in charge of bioinformatics data management. I.G., P.A.F., M.R.S. and P.J.C. coordinated the sequencing efforts and performed primary data analysis. X.S.P., X.E., A.L-G., C.L.-O. and E.C. directed the research and wrote the manuscript, which all authors have approved.

Author Information Sequencing, expression and genotyping array data have been deposited at the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/), which is hosted at the European Bioinformatics Institute (EBI), under accession number EGAS00000000092. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.C. (ecampo@clinic.ub.es) or C.L.-O. (clo@uniovi.es).

LETTER

doi:10.1038/nature10138

CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis

Bin-Zhi Qian[1], Jiufeng Li[1], Hui Zhang[1], Takanori Kitamura[1], Jinghang Zhang[2], Liam R. Campion[3], Elizabeth A. Kaiser[3], Linda A. Snyder[3] & Jeffrey W. Pollard[1]

Macrophages, which are abundant in the tumour microenvironment, enhance malignancy[1]. At metastatic sites, a distinct population of metastasis-associated macrophages promotes the extravasation, seeding and persistent growth of tumour cells[2]. Here we define the origin of these macrophages by showing that Gr1-positive inflammatory monocytes are preferentially recruited to pulmonary metastases but not to primary mammary tumours in mice. This process also occurs for human inflammatory monocytes in pulmonary metastases of human breast cancer cells. The recruitment of these inflammatory monocytes, which express CCR2 (the receptor for chemokine CCL2), as well as the subsequent recruitment of metastasis-associated macrophages and their interaction with metastasizing tumour cells, is dependent on CCL2 synthesized by both the tumour and the stroma. Inhibition of CCL2–CCR2 signalling blocks the recruitment of inflammatory monocytes, inhibits metastasis *in vivo* and prolongs the survival of tumour-bearing mice. Depletion of tumour-cell-derived CCL2 also inhibits metastatic seeding. Inflammatory monocytes promote the extravasation of tumour cells in a process that requires monocyte-derived vascular endothelial growth factor. *CCL2* expression and macrophage infiltration are correlated with poor prognosis and metastatic disease in human breast cancer[3–6]. Our data provide the mechanistic link between these two clinical associations and indicate new therapeutic targets for treating metastatic breast cancer.

To understand the origin of macrophages in primary tumours and their metastatic sites, we measured monocyte trafficking. Mouse monocytes were identified by their expression of CD11b and CD115 (Supplementary Fig. 3a) and were sorted by fluorescence-activated cell sorting (FACS) into sub-populations of inflammatory monocytes expressing Gr1 and Ly6c and resident monocytes lacking Gr1 and Ly6c (refs 7, 8) (Supplementary Fig. 3b–d). Both populations had similar expression of GFP in *Csf1r–GFP* transgenic mice (Supplementary Fig. 3b). We adoptively transferred[9–11] 10^5 cells of each population into syngeneic FVB mice bearing autochthonous late-stage Polyoma Middle T (PyMT) mammary tumours with spontaneous pulmonary metastases (Fig. 1a). Eighteen hours after adoptive transfer, we determined the ratio of recovered GFP-positive inflammatory monocytes (Supplementary Fig. 3e) to resident monocytes from the same donor, to measure their relative recruitment. This indicated that there were similar numbers of donor cells in the blood (showing equivalent availability), but that in the primary tumour, resident monocytes were preferentially recruited, whereas in pulmonary metastases, inflammatory monocytes were preferentially recruited, with more than threefold enrichment (Fig. 1b). Consistent with this, a notable population of endogenous inflammatory monocytes was identified in metastasis-bearing lungs but not in normal lungs (Supplementary Fig. 4a). This preferential recruitment of inflammatory monocytes in the lung was not observed in 7-week-old PyMT mice bearing pre-metastatic mammary tumours (Supplementary Fig. 4b). In experimentally induced pulmonary foci of intravenously injected Met-1 cells (a PyMT-induced mouse mammary tumour cell line)[12], inflammatory monocytes were also preferentially recruited (Supplementary Fig. 4c). GFP-labelled cells were readily detectable in pulmonary metastases at least 5 d after transfer (data not shown) and within 2 d, a significant portion of them had differentiated into F4/80$^+$ CD11b$^+$ Gr1$^-$ metastasis-associated macrophages (MAMs)[2] that are not seen in normal lungs (Supplementary Fig. 4d). To test whether inflammatory monocytes were recruited early in the metastasis process, we transferred monocyte populations

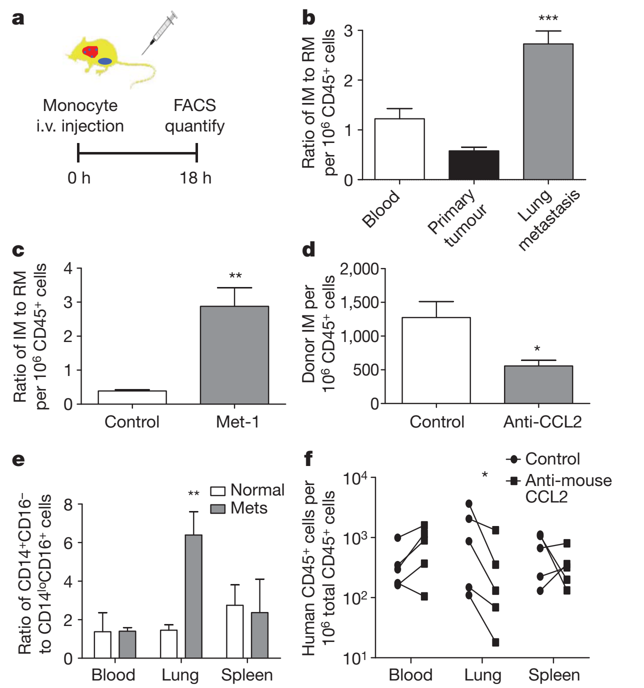


Figure 1 | Pulmonary metastases preferentially recruit inflammatory monocytes through CCL2. **a**, Schematic for the adoptive transfer of monocytes into PyMT-tumour-bearing mice with pulmonary metastases. i.v., intravenous. **b**, Ratios of inflammatory monocytes (IM) to resident monocytes (RM) in different tissues of recipient mice bearing PyMT tumours and metastases. $n = 6$; ***, $P < 0.0001$. **c**, Ratios of inflammatory monocytes to resident monocytes in control lungs and in lungs with Met-1 cells intravenously injected 7 h before measurement. $n = 4$; **, $P = 0.0039$. **d**, Relative numbers of donor inflammatory monocytes recruited in lungs challenged with Met-1 cells for 7 h, with control or anti-mouse CCL2 antibody treatment. $n = 3$; *, $P = 0.045$. **e**, Ratios of adoptively transferred CD14$^+$CD16$^-$ and CD14^{low}CD16$^+$ human monocytes recruited into the lungs of normal mice (open bars) and of mice challenged with 4173 cells that contain metastases (Mets: solid bars) for 7 h. $n = 5$; **, $P = 0.0163$. All bars show mean + s.e.m. **f**, Numbers of adoptively transferred human CD14$^+$CD16$^-$ monocytes that migrated into different tissues of mice challenged with 4173 cells via intravenous injection, with control or anti-mouse CCL2 antibody treatment. Each line connects data from the same donor. $n = 5$; *, $P = 0.016$.

[1]Department of Developmental and Molecular Biology, Center for the Study of Reproductive Biology and Women's Health, Albert Einstein College of Medicine, New York, New York 10461, USA. [2]Flow Cytometry Core Facility, Albert Einstein College of Medicine, New York, New York 10461, USA. [3]Ortho Biotech Oncology R&D, 145 King of Prussia Road, Radnor, Pennsylvania 19087, USA.

7 h after intravenous injection of Met-1 cells, a time point before significant interaction between the tumour and macrophages, and before extravasation of tumour cells[2]. Compared to control lungs, the recruitment of inflammatory monocytes to tumour-cell-challenged lungs increased markedly, with the ratio of inflammatory monocytes to resident monocytes increasing more than fivefold (Fig. 1c). However, this preferential recruitment of inflammatory monocytes was not observed after intravenous injection with PBS or latex beads, as controls for injection and particle lodgement, respectively (Supplementary Fig. 4e and data not shown). Consistent with this early recruitment of inflammatory monocytes, MAMs expressing high levels of CCR2 were preferentially recruited to lungs 36 h after tumour-cell inoculation[2]. However, B cells and T cells, including Foxp3^{+} regulatory T (T_{reg}) cells, were not differentially recruited at this time (Supplementary Fig. 5a–c and data not shown). These data indicate that MAMs are derived from inflammatory monocytes that are specifically recruited early in the process of pulmonary metastasis, before other immune cells.

Distinct chemokine signals recruit inflammatory and resident monocytes[7], with inflammatory monocytes responding to CCL2 (refs 10, 11). Lung metastases of PyMT tumours express CCL2 homogeneously, in contrast to its heterogeneous expression in primary tumours (Supplementary Fig. 6a–d), and inflammatory monocytes have high levels of CCR2 expression, whereas resident monocytes do not (Supplementary Fig. 6e). The neutralization of CCL2 using a CCL2-specific antibody[13] markedly inhibited both the recruitment of inflammatory monocytes to lungs challenged with metastatic tumour cells (Fig. 1d) and the increase in the number of MAMs at the metastatic site (Supplementary Fig. 4f). Other CCR2-expressing leukocytes (a sub-population of T cells) and also T_{reg} cells were unaffected by anti-CCL2 antibody treatment in this model (Supplementary Fig. 5d). Furthermore, the preferential recruitment of inflammatory monocytes to the tumour-cell-challenged lung was completely abrogated during adoptive transfer of monocytes sorted from *Ccr2*-null mice (Supplementary Fig. 6f).

The pattern of human monocyte recruitment to tumours *in vivo* is unknown. To investigate this, human CD14^{+}CD16^{-} inflammatory monocytes and CD14^{low}CD16^{+} resident monocytes[9] were sorted from enriched CD14^{+} cells from the peripheral blood of healthy donors (Supplementary Fig. 7a). 10^{5} cells of each population were adoptively transferred into pairs of nude mice supplemented with recombinant human colony-stimulating factor 1 (CSF1), which is essential for the survival of monocytes and macrophages (Supplementary Fig. 7e). Human monocytes were quantified 18 h after adoptive transfer, using FACS analysis with an antibody against human CD45 (Supplementary Fig. 7b). In normal mice, after adoptive transfer of monocytes from the same donor, there were comparable numbers of human inflammatory monocytes and resident monocytes in the circulation and also recruited to the lung, but about twice the numbers of inflammatory monocytes compared to resident monocytes in the spleen (Fig. 1e, open bars). In mice given an intravenous injection of human MDA-MB-231-derived metastatic 4173 breast cancer cells[14] 7 h before monocyte transfer, the ratio of the two monocyte populations in blood and spleen was similar to that in normal mice, but the ratio of inflammatory monocytes to resident monocytes in the lungs increased more than sixfold (Fig. 1e). In established pulmonary metastases derived from orthotopically injected 4173 cells, inflammatory monocytes were also preferentially recruited, with a ratio fivefold higher than that in normal lungs (Supplementary Fig. 7d). Mouse inflammatory monocytes were also preferentially recruited to lungs challenged with 4173 cells (data not shown). Human inflammatory monocytes express CCR2, whereas resident monocytes express minimal levels of this receptor (Supplementary Fig. 7c). The neutralization of host CCL2 with an antibody against mouse CCL2 markedly reduced the recruitment of human inflammatory monocytes into lungs challenged with 4173 cells, without any change in the circulation or spleen (Fig. 1f). Treatment with an antibody specific to human CCL2 (ref. 15) also inhibited inflammatory monocyte recruitment (Supplementary Fig. 7f), indicating the importance of CCL2 from both the tumour and the target organ. This shows that human inflammatory monocytes respond to the same CCL2–CCR2 signalling as mouse cells for their specific recruitment during pulmonary metastasis.

To test the effect on metastatic potential of blocking the recruitment of inflammatory monocytes, we performed experimental metastasis assays with Met-1 cells in mice treated with anti-mouse CCL2 or with a control antibody shortly before the tumour-cell injection. Anti-CCL2 treatment reduced the total metastasis burden, owing to a markedly reduced number of metastasis nodules (Fig. 2a, b). An antibody specific to mouse CCL12, another ligand of mouse CCR2, had no effect on the metastasis of Met-1 cells (Supplementary Fig. 8). This indicates that the specific CCL2-mediated recruitment of inflammatory monocytes is critical for the pulmonary seeding of tumour cells.

Extravasation is a critical step for the metastatic seeding of tumour cells in the lung[2]. We used an intact-lung imaging system[16] to test the role of CCL2-recruited inflammatory monocytes in tumour-cell extravasation. *Csf1r–GFP* transgenic mice were injected intravenously with cyan fluorescent protein (CFP)-expressing Met-1 cells and analysed after 24 h. Quantification of three-dimensional reconstructed confocal images (Fig. 2c, d and Supplementary Movies 1 and 2) showed that the number of macrophages interacting directly with tumour cells was significantly reduced by anti-mouse CCL2 neutralizing antibody, compared with control antibody (Fig. 2e). Notably, tumour-cell extravasation was delayed and less efficient after the blocking of inflammatory monocytes (Fig. 2f). Tumour-cell extravasation involves crosstalk between tumour cells, endothelial cells, basement membrane and macrophages. In an *in vitro* trans-endothelial migration assay (Supplementary Fig. 9a)[17], the trans-endothelial migration of tumour cells was enhanced about fivefold by mouse bone-marrow-derived macrophages (BMDMs) located on the basolateral side of the endothelial monolayer. This effect was blocked by anti-mouse-CCL2 neutralizing antibody, but not by control antibody (Supplementary Fig. 9b). Tumour cells, BMDMs and endothelial cells all express CCL2, whereas only the macrophages express CCR2 (Supplementary Fig. 9c), indicating that only macrophages respond to the CCL2 chemokine signalling. In confirmation of this, macrophages from *Ccr2*-null mice were not capable of promoting trans-endothelial migration of tumour cells (Supplementary Fig. 9d). Notably, FACS-sorted inflammatory monocytes, but not resident monocytes, markedly promoted tumour-cell trans-endothelial migration and this was also inhibited by anti-mouse-CCL2 neutralizing antibody (Fig. 2g, h).

Total blockade of CCL2 (both mouse and human) inhibited spontaneous lung metastasis of orthotopically injected MDA-MB-231 cells (Fig. 3a). Ligands secreted by both the tumour cells and the host contributed to metastatic efficiency, because both anti-human and anti-mouse antibodies markedly inhibited the experimental metastasis of 4173 cells (Fig. 3b) without affecting tumour-cell proliferation *in vitro* (data not shown). This conclusion was also confirmed by knocking down *CCL2* using small interfering RNAs in 4173 cells: this markedly reduced lung colonization in experimental metastasis assays (Supplementary Fig. 10e, f). Consistent with this, a similar *Ccl2*-knockdown in Met-1 cells did not affect tumour-cell proliferation *in vitro*, but markedly inhibited the metastatic efficiency of the cells (Supplementary Fig. 10a–c). Trans-endothelial migration of 4173 cells *in vitro* was also promoted by human inflammatory monocytes and inhibited by neutralizing either human or mouse CCL2 with specific antibodies (Fig. 3c–e). These data indicate that CCL2 secreted by both the tumour cell and the target organ promotes tumour-cell extravasation and metastatic seeding via the recruitment of inflammatory monocytes. Consistent with the role of CCL2 synthesized by the microenvironment in the lung, bone metastases of MDA-MB-231 cells also recruit inflammatory monocytes and the inhibition of CCL2 inhibits metastatic progression. In contrast, liver metastases of Met-1 cells did not recruit inflammatory monocytes and CCL2 inhibition did not reduce metastasis (data not shown). Furthermore, CCL2 blockade 2 d after intravenous injection of MDA-MB-231 cells reduced the tumour burden in the lung and prolonged the

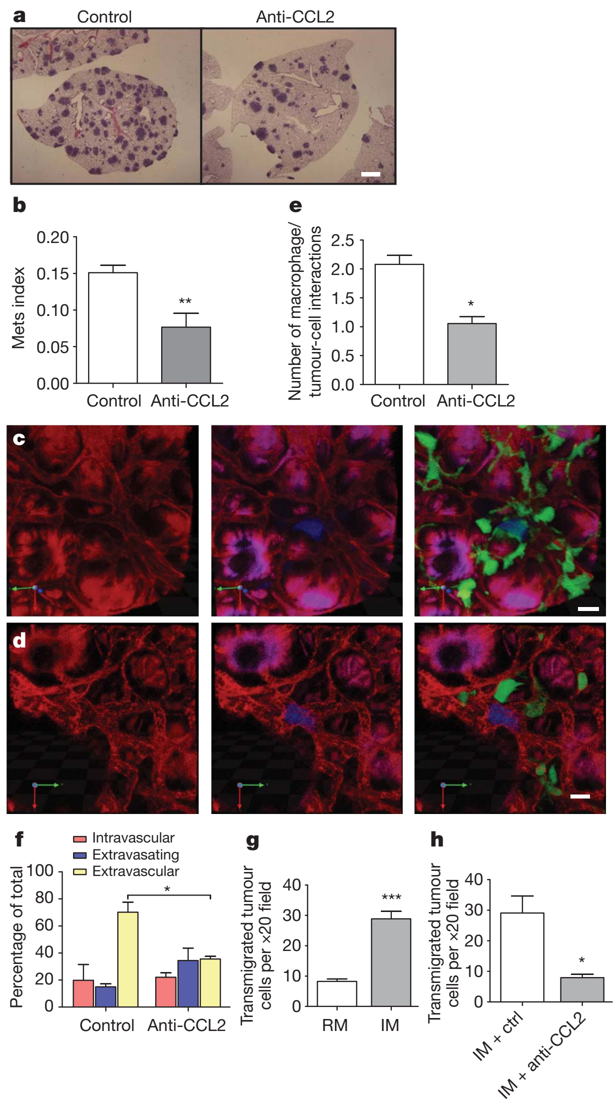


Figure 2 | CCL2-recruited monocytes promote metastatic seeding. **a**, Representative haematoxylin-&-eosin-stained sections showing Met-1 metastasis with control or anti-CCL2 antibody treatment. Scale bar, 1 mm. **b**, Met-1 metastasis (Mets) burden with or without antibody treatment. $n = 6$; **, $P = 0.006$. **c**, **d**, Representative snapshots of three-dimensional reconstructed confocal images of tumour cells (blue) and macrophages (green) in lung vasculature (red) 24 h after tail-vein injection of tumour cells into mice treated with control (**c**) or anti-mouse CCL2 (**d**) antibodies. Scale bar, 20 μm. Arrows define the dimensions of the figure. **e**, **f**, Numbers of interactions between macrophages and tumour cells (**e**) and tumour-cell extravasation (**f**) in mice with control or anti-mouse CCL2 antibody treatment. (**e**, $P = 0.0066$, and **f**, $P = 0.00163$, are based upon three-dimensional images of 15–20 tumour clusters per mouse, $n = 3$ mice per group.) **g**, Numbers of transmigrated Met-1 cells in the presence of resident monocytes or inflammatory monocytes. $n = 5$; ***, $P < 0.0001$. **h**, Numbers of transmigrated Met-1 cells in the presence of inflammatory monocytes, with antibody treatments. $n = 3$; *, $P = 0.0204$. All bars show mean + s.e.m.

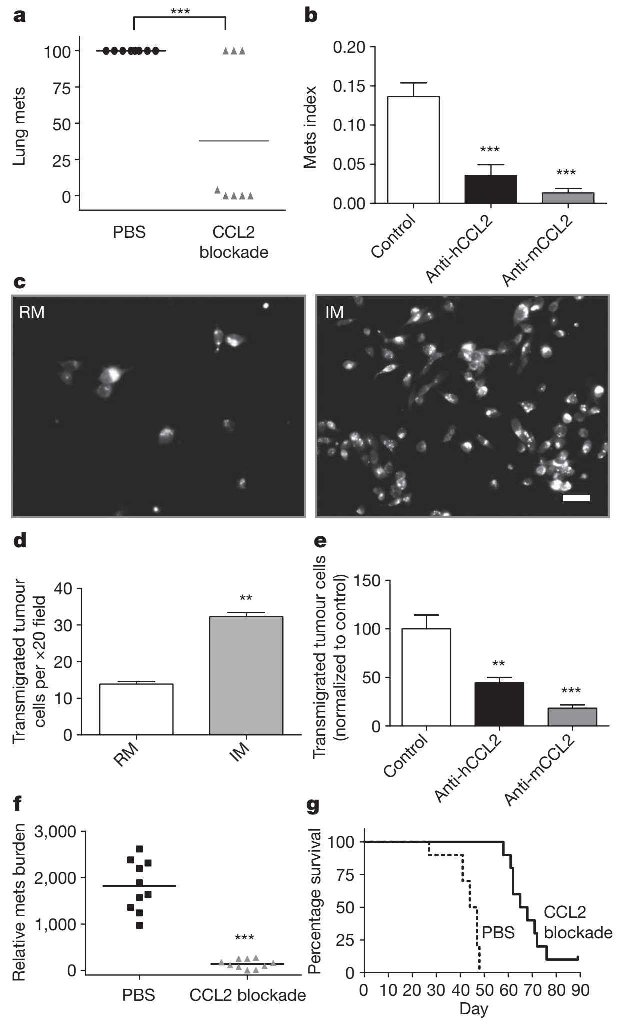


Figure 3 | CCL2 from both the tumour cell and the host promotes metastatic seeding. **a**, Numbers of spontaneous pulmonary metastases from orthotopic MDA-MB-231 tumours with total CCL2 blockade or control treatment. Bar shows the mean; $n = 8$ per group; ***, $P < 0.001$. **b**, Metastasis burden of intravenously injected 4173 cells with different antibody treatments. Bars show mean + s.e.m.; $n = 6$; ***, $P = 2.14 \times 10^{-5}$. **c**, Representative fluorescent micrographs of transmigrated human 4173 cells pre-stained with cell-tracker dye in the presence of inflammatory or resident monocytes. Scale bar, 20 μm. **d**, Numbers of transmigrated 4173 cells in the presence of inflammatory monocytes or resident monocytes. Bars show the mean + s.e.m. of three experiments with duplicates; **, $P = 0.0051$. **e**, Relative number of transmigrated 4173 cells in the presence of inflammatory monocytes with control, anti-human CCL2 or anti-mouse CCL2 antibodies, normalized to the average number with control antibody treatment, which is set to 100. Bars represent the mean + s.e.m. of five experiments with duplicates. One-way analysis of variance with Bonferroni's multiple comparison test; **, $P < 0.01$; ***, $P < 0.001$. **f**, **g**, CCL2 blockade starting from 2 d after intravenous injection of MDA-MB-231 cells significantly reduces the metastasis burden, as measured by real-time PCR of human Alu repeats, normalized to mouse β-actin, on day 22 (**f**, $n = 10$; ***, $P < 0.001$). CCL2 blockade also prolongs survival (**g**, $n = 10$, $P < 0.001$) compared to control treatment with PBS.

survival of mice, indicating the importance of continuous recruitment of inflammatory monocytes and their differentiation into MAMs for persistent metastatic growth (Fig. 3f, g).

To determine a mechanism for the effects of inflammatory monocytes on tumour-cell extravasation, we analysed the transcriptomes of resident and inflammatory monocytes[18]. Among the differentially regulated genes, vascular endothelial growth factor A (*Vegfa*) was highly expressed by inflammatory monocytes, a fact that we verified experimentally (Supplementary Fig. 11a). To ablate *Vegfa* conditionally in myeloid cells to test its role in the metastatic process, we generated a transgenic mouse expressing a tamoxifen-inducible Mer-iCre fusion protein driven by the *Csf1r* promoter, crossed with *Vegfa*$^{\text{flox/flox}}$ mice[19]. Inducible ablation of *Vegfa* was achieved in cultured BMDMs treated with 4-hydroxytamoxifen (Fig. 4a) and these *Vegfa*-null BMDMs were unable to promote the trans-endothelial migration of tumour cells and did not enhance permeability of the endothelial monolayer, a process important for metastasis[20], when compared to control BMDMs (Fig. 4b, c). *In vivo* injection of tamoxifen specifically ablated *Vegfa* in monocytes, without ablation in other circulating immune cells (Fig. 4d). This monocyte-specific depletion of VEGFA markedly inhibited the potential for experimental

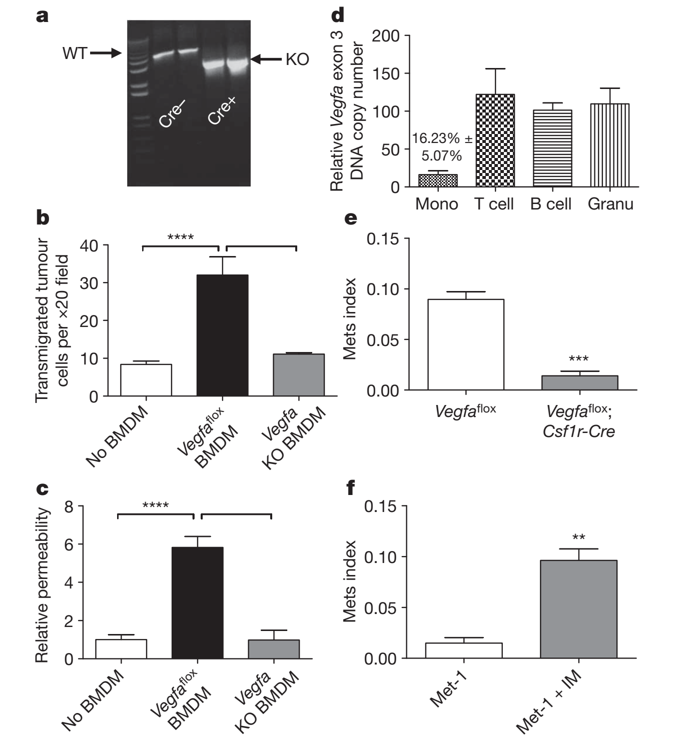


Figure 4 | Monocyte-specific ablation of *Vegfa* blocks pulmonary seeding. **a**, PCR of *Vegfa* exon 3 in BMDMs from *Vegfa*$^{flox/flox}$ mice, with or without the *Csf1r-Mer-iCre-Mer* transgene, treated with 4-hydroxytamoxifen. Wild-type (WT) and knockout (KO) bands are indicated. **b**, **c**, Numbers of trans-endothelial migrated Met-1 cells (**b**) and permeability of the endothelial monolayer to albumin (**c**), with no BMDMs, *Vegfa*^{flox} BMDMs or *Vegfa*-knockout BMDMs. $n = 3$ with duplicates; **, $P < 0.01$ with analysis of variance. **d**, Relative copy number of *Vegfa* exon 3 in leukocytes from the peripheral blood of tamoxifen-treated *Vegfa*$^{flox/flox}$ *Csf1r-Mer-iCre-Mer* mice compared with *Vegfa*$^{flox/flox}$ mice. Mono, monocyte; granu, granulocyte. **e**, Met-1 Mets burden in *Vegfa*$^{flox/flox}$ mice with or without *Cre*, with the same tamoxifen treatment. $n = 6$; ***, $P = 0.0004$. **f**, Met-1 Mets burden in *Vegfa*$^{flox/flox}$ *Csf1r-Mer-iCre-Mer* mice with tamoxifen treatment, with or without co-injection of inflammatory monocytes. $n = 6$; **, $P < 0.0001$. All data are mean + s.e.m.

metastasis of Met-1 cells and reduced their seeding efficiency (Fig. 4e and Supplementary Fig. 11b). Adoptive transfer experiments indicated that *Vegfa*-null inflammatory monocytes infiltrate Met-1 lung metastases at a comparable level to *Vegfa*^{flox} inflammatory monocytes, showing that VEGFA is not required for the recruitment of these cells (Supplementary Fig. 11c). Notably, co-injection of Met-1 cells and wild-type inflammatory monocytes into inducible macrophage-*Vegfa*-knockout mice restored the metastatic potential of tumour cells (Fig. 4f).

These experiments indicate that CCL2 synthesized by metastatic tumour cells and by the target-site tissue stroma is critical for the recruitment of a sub-population of CCR2-expressing monocytes that enhance the subsequent extravasation of the tumour cells. Mechanistically, this occurs at least in part through targeted delivery of molecules such as VEGFA that promote extravasation. Inflammatory monocytes are continually recruited by a CCL2-dependent mechanism and differentiate into macrophages that promote the subsequent growth of metastatic cells (Supplementary Fig. 1). These data, together with the clinical association of CCL2 overexpression in human cancers with poor prognosis (Supplementary Fig. 2), strongly argue for therapeutic approaches targeted against monocyte recruitment and function.

METHODS SUMMARY

The trafficking of monocytes into primary tumours and their metastases was studied by adoptive transfer of mouse (Ly6c/Gr1^{+} or Ly6c/Gr1^{-}) monocytes or human (CD14^{+}CD16^{+} and CD16^{-}) monocytes, using MMTV-PyMT autochthonous, human and mouse experimental metastasis models and human orthotopic tumour models. Monocytes and macrophages were recovered by enzymatic disaggregation of the tumours, followed by FACS analysis. To investigate mechanisms for monocyte recruitment and the effect of inhibition of this recruitment on metastasis, anti-mouse-CCL2 or anti-human-CCL2 antibodies or *Ccr2*-null mutant mice were used. To ablate *Vegfa* expression in monocytes, a myeloid-specific (*Csf1r* promoter), tamoxifen-inducible Cre-expressing strain was crossed with *Vegfa*$^{flox/flox}$ mice and gene ablation was induced by tamoxifen. The effect of monocyte depletion on tumour-cell extravasation using Met-1, an FVB PyMT-tumour-derived metastatic cell line, was determined using an *ex vivo* intact-lung imaging system and an *in vitro* extravasation assay.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 May 2010; accepted 19 April 2011.
Published online 8 June 2011.

1. Qian, B. Z. & Pollard, J. W. Macrophage diversity enhances tumor progression and metastasis. Cell **141,** 39–51 (2010).
2. Qian, B. *et al.* A distinct macrophage population mediates metastatic breast cancer cell extravasation, establishment and growth. *PLoS ONE* **4,** e6562 (2009).
3. Ueno, T. *et al.* Significance of macrophage chemoattractant protein-1 in macrophage recruitment, angiogenesis, and survival in human breast cancer. *Clin. Cancer Res.* **6,** 3282–3289 (2000).
4. Valkovic, T., Lucin, K., Krstulja, M., Dobi-Babic, R. & Jonjic, N. Expression of monocyte chemotactic protein-1 in human invasive ductal breast cancer. *Pathol. Res. Pract.* **194,** 335–340 (1998).
5. Saji, H. *et al.* Significant correlation of monocyte chemoattractant protein-1 expression with neovascularization and progression of breast carcinoma. *Cancer* **92,** 1085–1091 (2001).
6. Rhodes, D. R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6,** 1–6 (2004).
7. Geissmann, F. *et al.* Blood monocytes: distinct subsets, how they relate to dendritic cells, and their possible roles in the regulation of T-cell responses. *Immunol. Cell Biol.* **86,** 398–408 (2008).
8. Geissmann, F. *et al.* Development of monocytes, macrophages, and dendritic cells. *Science* **327,** 656–661 (2010).
9. Geissmann, F., Jung, S. & Littman, D. R. Blood monocytes consist of two principal subsets with distinct migratory properties. *Immunity* **19,** 71–82 (2003).
10. Getts, D. R. *et al.* Ly6c^{+} "inflammatory monocytes" are microglial precursors recruited in a pathogenic manner in West Nile virus encephalitis. *J. Exp. Med.* **205,** 2319–2337 (2008).
11. Palframan, R. T. *et al.* Inflammatory chemokine transport and presentation in HEV: a remote control mechanism for monocyte recruitment to lymph nodes in inflamed tissues. *J. Exp. Med.* **194,** 1361–1374 (2001).
12. Borowsky, A. D. *et al.* Syngeneic mouse mammary carcinoma cell lines: two closely related cell lines with divergent metastatic behavior. *Clin. Exp. Metastasis* **22,** 47–59 (2005).
13. Tsui, P. *et al.* Generation, characterization and biological activity of CCL2 (MCP-1/JE) and CCL12 (MCP-5) specific antibodies. *Hum. Antibodies* **16,** 117–125 (2007).
14. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436,** 518–524 (2005).
15. Carton, J. M. *et al.* Codon engineering for improved antibody expression in mammalian cells. *Protein Expr. Purif.* **55,** 279–286 (2007).
16. Al-Mehdi, A. B. *et al.* Intravascular origin of metastasis from the proliferation of endothelium-attached tumor cells: a new model for metastasis. *Nature Med.* **6,** 100–102 (2000).
17. Ma, C. & Wang, X.-F. *In vitro* assays for the extracellular matrix protein-regulated extravasation process. *CSH Protoc.* doi:10.1101/pdb.prot5034 (2008).
18. Swirski, F. K. *et al.* Identification of splenic reservoir monocytes and their deployment to inflammatory sites. *Science* **325,** 612–616 (2009).
19. Gerber, H. P. *et al.* VEGF is required for growth and survival in neonatal mice. *Development* **126,** 1149–1159 (1999).
20. Huang, Y. *et al.* Pulmonary vascular destabilization in the premetastatic phase facilitates lung metastasis. *Cancer Res.* **69,** 7529–7537 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the NIH to J.W.P. (NIH PO1 CA100324 and RO1 CA131270) and to the Albert Einstein Cancer Center Core (P30 CA 13330). We thank J. Massague for 4173 cells and N. Ferrara for the *Vegfa*$^{flox/flox}$ mice. We also thank P. Marsters for statistical analyses and M. Thompson, F. Shi, C. Ferrante, F. McCabe, H. Millar-Quinn and D. Wiley for discussions and technical assistance.

Author Contributions B-Z.Q., L.A.S. and J.W.P. conceived the ideas and designed the experiments. B-Z.Q., J.L., H.Z., T.K., J.Z., L.R.C. and E.A.K. performed the experiments. B-Z.Q., J.L., L.A.S. and J.W.P. analysed the data. B-Z.Q., L.A.S. and J.W.P. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.W.P. (pollard@aecom.yu.edu).

METHODS

Animals. All procedures involving mice were conducted in accordance with National Institutes of Health regulations concerning the care and use of experimental animals. The study of mice was approved by the Albert Einstein College of Medicine and Ortho Biotech R&D Institute animal care and use committees. Transgenic mice expressing the Polyoma Middle T (PyMT) oncogene under the control of the mouse mammary tumour virus long terminal repeat (MMTV LTR) promoter were provided by W. J. Muller and were bred in-house. FVB (*Tg(Csf1r-EGFP)1Jwp*) mice have been previously reported to have the whole mononuclear phagocyte system labelled[2]. BL6 *Ccr2*$^{\mathrm{tm1Ifc}}$/J mice were purchased from The Jackson Laboratory. The FVB macrophage-specific (*Csf1r* promoter), tamoxifen-inducible Cre-expressing *Tg*(*Csf1r-Mer-iCre-Mer*)*1Jwp* transgenic mouse strain was generated and crossed with *Vegfa*$^{\mathrm{flox/flox}}$ mice (gift from N. Ferrara). Knockout of *Vegfa* in myeloid cells was induced by daily subcutaneous injection of 3 μg tamoxifen per mouse for 2 d, before sorting for blood leukocytes or tumour-cell injection.

Metastasis assay. Eight-week-old FVB females and six-week-old female nude mice were used for lung experimental-metastasis assays with intravenous injection of 5×10^5 Met-1 cells or 10^6 MDA-MB-231-derived LM2 human breast cancer cells, 4173 (ref. 14), respectively. If not otherwise specified, all animals were killed 2 weeks after intravenous injection of Met-1 cells or 4 weeks after injection of human tumour cells, for optimal metastatic burden. In experimental metastasis assays, antibodies were given at 10 mg kg^{-1} body weight via intraperitoneal injection 3 h before tumour-cell injection, for single treatments, or twice a week thereafter for prolonged treatments, if not otherwise specified. For paraffin sections, lungs were injected with 1.2 ml of 10% neutral buffered formalin by tracheal cannulation to fix the inner airspaces and inflate the lung lobes. Lungs were excised and fixed in formalin overnight. A precise stereological method[21] with modification was used for quantification of lung metastases. Briefly, paraffin-embedded lungs were systematically sectioned through the entire lung with one 5 μm section taken in every 0.5 mm of lung thickness. All sections were stained with haematoxylin and eosin and images were taken using a Zeiss SV11 microscope with a Retiga 1300 digital camera and analysed using ImageJ[22]. The Mets index is the total volume of metastases normalized to total lung volume and Mets number is the number of metastasis nodules per mm^2 of lung area. Real-time PCR quantification of the burden of human tumour cells was performed as reported previously, using human-specific primers[23]. For spontaneous lung metastasis, 2.5×10^6 parental MDA-MB-231 cells or 10^6 derived LM2 4173 tumour cells were orthotopically injected into the inguinal mammary gland of SCID beige or nude mice, respectively. Anti-mouse CCL2 and CCL12 (ref. 23) and anti-human CCL2 (ref. 15) antibodies neutralize only their respective target molecules and were provided by Ortho Biotech Oncology together with the control antibody. In spontaneous metastasis assays, antibody treatment began on day 3 after tumour-cell intra-mammary-gland injection and continued twice a week thereafter, with each antibody used at 20 mg kg^{-1} body weight. When each group reached a mean primary-tumour volume of ~1,000 mm^3, the mice were killed. Lungs were perfused with India ink and placed in Fekete's solution. Lung metastases were counted in a blinded fashion. All *in vivo* experiments were at least two independent experiments with 3–10 mice for each group.

Adoptive transfer. CD115$^+$ F4/80$^+$ CD11b$^+$ Ly6c1/Gr1$^+$ and Ly6c1/Gr1$^-$ bone-marrow monocytes were sorted from FVB Csf1r-EGFP mice and adoptively transferred into FVB mice. 10^5 of either cell type were transferred into mice bearing mammary tumours and/or pulmonary metastases. Monocytes were sorted from *Ccr2*-null mutant mice using the same protocol and were labelled with CellTracker (Invitrogen) following the manufacturer's instructions, before adoptive transfer into nude mice. Fresh human CD14$^+$ peripheral monocytes were purchased from All Cells LLC. 10^5 CD14$^+$CD16$^-$ and CD14$^+$CD16$^+$ cells were FACS-sorted and intravenously transferred into nude mice supplemented with 2×10^6 units of recombinant human CSF1 via subcutaneous injection. In the indicated experiments, specified antibodies were given at 10 mg kg^{-1} body weight 3 h before adoptive transfer of monocytes.

FACS analysis and antibodies. For FACS analysis, lungs or whole mice were perfused thoroughly with cold PBS before cell collection, then lungs were minced on ice and digested with an enzyme mix of Liberase and Dispase (Invitrogen). Blood was drawn by cardiac puncture. Red blood cells were removed using RBC lysis buffer (eBioscience). Cells were blocked using anti-mouse CD16/CD32 antibody (eBioscience) for mouse cells, or 10% goat serum for human cells, before antibody staining. Antibodies against mouse antigens were: CD45 (30-F11), CD11b (M1/70), Gr1 (RB6-8C5), CD115 (AFS98) and Foxp3 (FJK-16 s; all from eBioscience); CD3 (145-2C11) and Ly6c1 (HK1.4; both from Biolegend); CD25 (PC61), CD62L (MEL-14), IL4Ra (mIL4R-M1), CD4 (GK1.5), CD8a (53-6.7) and Ly6G (1A8; all from BD Pharmingen) and F4/80 (Cl:A3-1; AbD Serotec). Antibodies against human antigens were: CD14 (Tük4) and CD16 (3G8; both from Invitrogen), CD45 (HI30; BioLegend) and CCR2 (48607; R&D Systems). FACS analysis was performed on a LSRII cytometer (BD Biosciences) and data were analysed using Flowjo software (TreeStar). Gating of single cells using FSC/W and SSC/W and exclusion of dead cells with DAPI staining were performed routinely during analysis. Mouse CCL2 was stained using the specific antibody R-17 (Santa Cruz) after a standard immunohistochemistry protocol.

Cell culture and *in vitro* extravasation assay. All cells were cultured in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum (FBS). The extravasation assay was performed as previously described[17,24] with modifications. Briefly, 2×10^4 endothelial cells (3B-11, ATCC) were plated into the upper chamber of a GFR matrigel invasion chamber (BD Biosciences) in DMEM with 10% (v/v) FBS. A monolayer was formed in 2 d and was verified by microscopy. 10^4 BMDMs or FACS-sorted monocytes were loaded to the basolateral side of the insert and put into a plate-well with DMEM, 10% FBS and 10^4 units ml^{-1} CSF1 to allow attachment. *Vegfa*-knockout BMDMs derived from *Csf1r-Mer-iCre-Mer:Vegfa*$^{\mathrm{flox/flox}}$ mice were induced by treating the cells with 1 μM 4-hydroxyltamoxifen for 7 d after isolation of bone marrow. 2×10^4 Met-1 cells stained with CellTracker CMRA (Invitrogen) were loaded into the insert with DMEM in 0.5% (v/v) FBS and 10^4 units ml^{-1} CSF1. CCL2-neutralizing antibody and control antibody were used at 5 ug ml^{-1}, applied to both sides of the insert. Plates were incubated under normal tissue-culture conditions for 36–48 h before being fixed with 1% (w/v) paraformaldehyde. Tumour-cell trans-endothelial migration was quantified by counting the number of cells that migrated through the insert under a fluorescent microscope (6–10 randomly-selected fields in each insert) and was expressed as cell number per ×20 field, if not otherwise specified. The permeability assay was performed by loading 4% (w/v) bovine serum albumin labelled with Evan's blue into the upper chamber with a pre-formed endothelial monolayer of 3B-11 cells and measuring the absorption of the phenol-red-free medium in the lower chamber at 650 nm after a 30-min incubation in normal culturing conditions. All *in vitro* experiments were at least three independent experiments with duplicate or triplicate measures.

Molecular biology. To knockdown *Ccl2* in Met-1 cells, a 97-mer oligo containing a small hairpin RNA (shRNA) that targets the *Ccl2* mRNA sequence from nucleotide 166 was cloned into the miR30 context in the retroviral vector P2GM[25]. To knockdown *CCL2* in 4173 cells, a 97-mer oligo containing a shRNA targeting the human *CCL2* mRNA sequence from nucleotide 255 was cloned into the miR30 context in the same vector. For real-time PCR of mouse *Ccl2* expression, primers CCCAATGAGTAGGCTGGAGA and AAAATGGATCCACACCTTGC were used, and for *Ccr2*, primers CCTGCAAAGACCAGAAGAGG and GTGAGCAGGAAGAGCAGGTC. All real-time PCR was performed on an MJ Research DNA Engine 2 Opticon real-time PCR machine using SYBR master mix (Invitrogen). Primers used were: mouse *Ccl2* primers GTTGGCTCAGCCAGATGCA and AGCCTACTCATTGGGATCATCTTG; mouse *Ccr2* TTTGTTTTTGCAGATGATTCAA and TGCCATCATAAAGGAGCCAT; mouse *Plau* ACAGATAAGCGGTCCTCCAG and GCCCCACTACTATGGCTCTG; mouse *Vegfa* AATGCTTTCTCCGCTCTGAA and GCTTCCTACAGCACAGCAGA; mouse *Vegfa* exon 3 ACATCTTCAAGCCGTCCTGT and CTGCATGGTGATGTTGCTCT; human *CCL2* AGGTGACTGGGGCATTGAT and GCCTCCAGCATGAAAGTCTC. To verify mouse *Vegfa* exon 3 knockout, primers that flank this exon, GCTGCACCCACGACAGAAGG and TGAGGTTTGATCCGCATGAT, were used.

***Ex vivo* whole-lung imaging.** A well-established intact-lung microscopy technique[16,26] was applied to observe tumour cells, macrophages and blood vessels in mouse lungs. CFP-expressing Met-1 cells, prepared by retrovirus infection of a CMV-promoter CFP vector, were injected intravenously into the tail vein of each mouse. At the times indicated, mice were anaesthetized and injected with 10 μg AlexaFluor-647-conjugated anti-mouse CD31 antibody (BioLegend). Five minutes later, the mouse was put under artificial ventilation through tracheal cannulation. The lung was cleared of blood by gravity perfusion through the pulmonary artery with artificial medium (Kreb-Ringer bicarbonate buffer with 5% dextran and 10 mmol l^{-1} glucose (pH 7.4)). The heart–lung preparation was dissected en bloc and placed in a specially designed plexiglass chamber with a port to the artificial cannula. The lung rested on a plexiglass window at the bottom of the chamber with the posterior surface of the lung touching the plexiglass. The lung was ventilated throughout the experiment with 5% CO_2 in medical air and perfused by gravity perfusion except during imaging. Three to five animals were imaged for each time point and 10–20 unrelated fields were imaged for each animal.

Images were collected with a Leica TCS SP2 AOBS confocal microscope (Mannheim) with ×60 oil-immersion optics. Laser lines at 458 nm, 488 nm and 633 nm for excitation of CFP, GFP and AF647, respectively, were provided by an Ar laser and a HeNe laser. Detection ranges were set to eliminate crosstalk between fluorophores. Three-dimensional reconstruction was performed using Volocity (Improvision Inc.).

Statistical analysis. Statistical analysis methods were the standard two-tailed Student's *t*-test for two data sets and ANOVA followed by Bonferroni/Dunn post

hoc tests for multiple data sets using Prism (GraphPad Inc.), except for human-monocyte transfer with antibody treatment, where a paired *t*-test was used because of variations among different donors. For the spontaneous-metastasis assay of MDA-MB-231 cells, percentage differences in numbers of lung metastases were compared between groups using parametric survival regression methods, with metastasis counts of more than 100 considered censored at 100. *P* values of less than 0.05 were deemed significant.

21. Nielsen, B. S. *et al.* A precise and efficient stereological method for determining murine lung metastasis volumes. *Am. J. Pathol.* **158,** 1997–2003 (2001).
22. Abramoff, M. D., Magelhaes, P. J. & Ram, S. J. Image processing with ImageJ. *Biophotonics Int.* **11,** 36–42 (2004).
23. Havens, A. M. *et al.* An *in vivo* mouse model for human prostate cancer metastasis. *Neoplasia* **10,** 371–380 (2008).
24. Brandt, B. *et al.* 3D-extravasation model — selection of highly motile and metastatic cancer cells. *Semin. Cancer Biol.* **15,** 387–395 (2005).
25. Stern, P. *et al.* A system for Cre-regulated RNA interference *in vivo*. *Proc. Natl Acad. Sci. USA* **105,** 13895–13900 (2008).
26. Im, J. H. *et al.* Coagulation facilitates tumor cell spreading in the pulmonary vasculature during early metastatic colony formation. *Cancer Res.* **64,** 8613–8619 (2004).

LETTER

doi:10.1038/nature10188

De novo cardiomyocytes from within the activated adult heart after injury

Nicola Smart[1]*, Sveva Bollini[1]*, Karina N. Dubé[1], Joaquim M. Vieira[1], Bin Zhou[2,3,4], Sean Davidson[5], Derek Yellon[5], Johannes Riegler[6,7], Anthony N. Price[8], Mark F. Lythgoe[6], William T. Pu[2,3] & Paul R. Riley[1]

A significant bottleneck in cardiovascular regenerative medicine is the identification of a viable source of stem/progenitor cells that could contribute new muscle after ischaemic heart disease and acute myocardial infarction[1]. A therapeutic ideal—relative to cell transplantation—would be to stimulate a resident source, thus avoiding the caveats of limited graft survival, restricted homing to the site of injury and host immune rejection. Here we demonstrate in mice that the adult heart contains a resident stem or progenitor cell population, which has the potential to contribute bona fide terminally differentiated cardiomyocytes after myocardial infarction. We reveal a novel genetic label of the activated adult progenitors via re-expression of a key embryonic epicardial gene, Wilm's tumour 1 (*Wt1*), through priming by thymosin β4, a peptide previously shown to restore vascular potential to adult epicardium-derived progenitor cells[2] with injury. Cumulative evidence indicates an epicardial origin of the progenitor population, and embryonic reprogramming results in the mobilization of this population and concomitant differentiation to give rise to *de novo* cardiomyocytes. Cell transplantation confirmed a progenitor source and chromosome painting of labelled donor cells revealed transdifferentiation to a myocyte fate in the absence of cell fusion. Derived cardiomyocytes are shown here to structurally and functionally integrate with resident muscle; as such, stimulation of this adult progenitor pool represents a significant step towards resident-cell-based therapy in human ischaemic heart disease.

Two previous studies have indicated a significant contribution of embryonic epicardial progenitor cells (EPDCs) to the cardiomyocyte lineage[3,4]. We investigated a basis for translating this myocardial potential in the adult heart. A significant problem in this regard is the lack of current adult epicardium-specific markers and authentic adult EPDC–Cre-expressing mouse strains for canonical lineage tracing. Previously reported genetic models to trace, or target, cells originating in the epicardium, such as $Wt1^{GFPCre/+}$ and $Wt1^{CreERT2/+}$; $R26R^{EYFP/+}$ mice or $Tbx18^{Cre}$ mice[3,4], cannot be applied directly to their adult counterparts, as the epicardial markers are either restricted to embryonic stages (Supplementary Fig. 1) or additionally expressed in the myocardium, as in the case of the *Tbx18* model[5]. Therefore, we sought to reactivate *Wt1* expression in the adult heart by pre-treatment ('priming') with thymosin β4 (Tβ4), which we previously showed induces adult EPDCs to form vascular precursors for neovascularization[2,6], followed by myocardial infarction (see schematic, Fig. 1a). Thus, we were able to establish both constitutive (GFP^+) and pulse (YFP^+) labelling of $Wt1^+$ progenitors to characterize the potential spatiotemporal distribution of primed adult cardiomyocyte precursors.

We initially established epicardial explants from Tβ4-primed $Wt1^{GFPCre/+}$ adult hearts (7 days of intraperitoneal injections without myocardial infarction; Supplementary Fig. 2a–d) as previously described[2,7] and investigated Isl1 expression as a marker of postnatal cardioblasts[8] along with Nkx2-5, an early marker of cardiomyocyte progenitors[9–11]. $Isl1^+/GFP^+$ cells and $Nkx2\text{-}5^+/GFP^+$ cells were prevalent within the explant cultures with mean percentage incidences of $76.7 \pm 6.3\%$ and $7.2 \pm 2.1\%$, respectively (mean percentage ± standard error of mean (s.e.m.); $n = 12$ explants), as were progenitor-like GFP^+ cells, which expressed more mature markers of cardiomyocyte differentiation such as cardiac troponin T (cTnT, also known as Tnnt2; $4.1 \pm 1.6\%$; Supplementary Fig. 2i, j) sarcomeric α-actinin (SαA; $4.8 \pm 2.4\%$; Supplementary Fig. 2l, m) and cardiac myosin binding protein C (MyBPC, also known as Mybpc2; 4.6 ± 0.8, mean percentage ± s.e.m.; Supplementary Fig. 2o, p). By day 14 in culture, cells adopted a more differentiated cardiac muscle phenotype; with evidence of sarcomeric structure in conjunction with compartmentalization of the GFP signal (Supplementary Fig. 2k–r).

Next we determined the extent of *Wt1* re-expression *in vivo* alongside the quantity and distribution of GFP^+ and YFP^+ adult progenitors. With injury alone (no Tβ4 priming), expression of *Wt1* and *Tbx18* was significantly increased at day 7 after myocardial infarction, dependent on the severity of the injury (Supplementary Fig. 3a). Following Tβ4 priming, expression of both epicardial genes was precociously increased by day 2 after myocardial infarction (Supplementary Fig. 3b), and this persisted in Tβ4-primed GFP^+ cells isolated by fluorescence-activated cell sorting (FACS) at day 4 after injury (Supplementary Fig. 4a). *In situ* hybridization, at an equivalent stage, revealed an upregulation in *Wt1* expression in small round 'progenitor-like' cells within the epicardium, subepicardial region and underlying myocardium (Supplementary Fig. 4b–d).

Tβ4 priming resulted in significantly more sorted GFP^+ and YFP^+ cells from whole hearts taken at day 7 after myocardial infarction when compared to treatment with vehicle (Fig. 1b–e and Supplementary Fig. 5a–d; GFP, 6.12% +Tβ4 versus 3.38% +PBS; YFP, 0.74% +Tβ4 versus 0.36% +PBS). This was confirmed *in situ* by anti-GFP immunostaining on serial heart sections at the equivalent stage after injury (Supplementary Fig. 5e). Further flow cytometry characterization of a progenitor phenotype revealed that labelled cells, at day 4, were not $c\text{-}Kit^+$ (Supplementary Fig. 6a–e). Instead, approximately 80% were positive for stem cell antigen factor 1 ($Sca\text{-}1^+$; Supplementary Fig. 6f–j), consistent with the notion that the adult epicardium is a heterogeneous lineage[12].

Two-photon molecular excitation laser scanning microscopy revealed pulse-labelled YFP^+ cells in the epicardium and subepicardial region at day 7 after myocardial infarction (Fig. 1f, g), distributed in diminishing numbers towards the underlying myocardium (Fig. 1h). Proliferative $Ki67^+$ progenitors were observed in epicardial and subepicardial regions (Supplementary Fig. 7a, b) alongside YFP^+ cells positive for phospho-histone H3 (Supplementary Fig. 7c–f), which

[1]Molecular Medicine Unit, UCL Institute of Child Health, London WC1N 1EH, UK. [2]Harvard Stem Cell Institute and Department of Cardiology, Children's Hospital Boston, 300 Longwood Avenue, Boston, Massachusetts 02115, USA. [3]Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. [4]Institute for Nutritional Sciences, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, China, 20031. [5]The Hatter Cardiovascular Institute, University College London, London WC1E 6HX, USA. [6]Centre for Advanced Biomedical Imaging (CABI), Department of Medicine and Institute of Child Health, University College London, London WC1E 6DD, UK. [7]Centre for Mathematics and Physics in the Life Sciences and Experimental Biology (CoMPLEX), University College London, London WC1E 6BT, UK. [8]MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College London, London W12 0NN, UK.
*These authors contributed equally to this work.

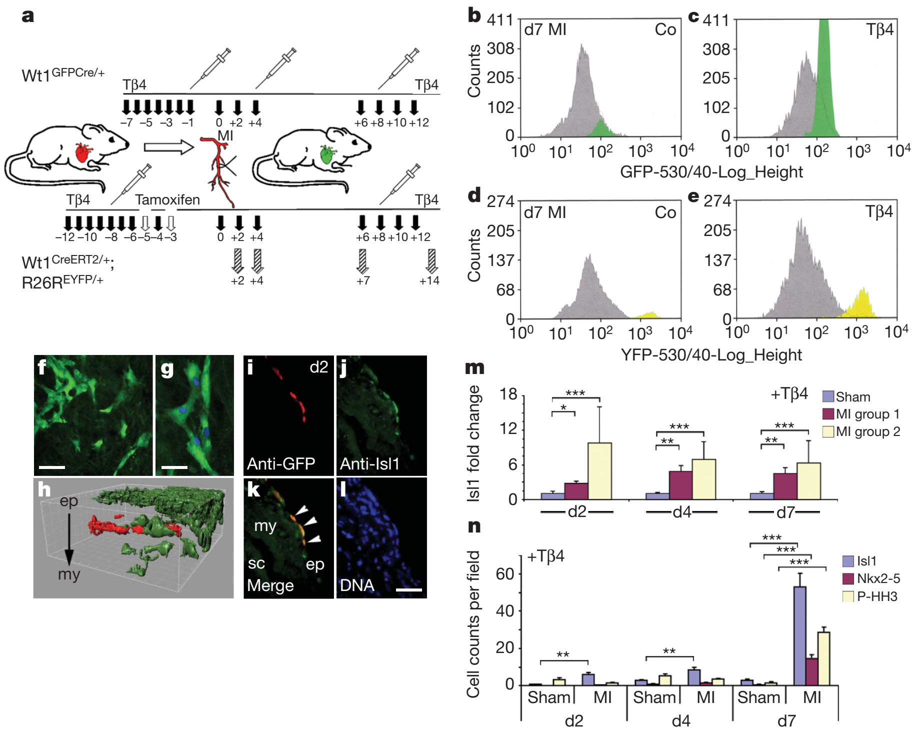


Figure 1 | Activated Wt1$^+$ cells give rise to cardiac progenitors in the injured adult heart. a, Schematic of constitutive or pulse-chase labelling of Wt1$^+$ cells. **b–e**, FACS analyses of whole hearts at day 7 (d7) after myocardial infarction (MI) revealed a significant increase in GFP$^+$ (**b**, **c**) and YFP$^+$ (**d**, **e**) cells following priming with Tβ4, as compared to PBS-treated controls (Co); *x*-axes represent either GFP (**b**, **c**) or YFP (**d**, **e**) fluorescent wavelengths on a logarithmic scale and *y*-axes represent total cell numbers isolated by FACs (**b–e**). **f**, **g**, Multi-photon imaging at day 7 after myocardial infarction revealed YFP$^+$ cells within the epicardium and subepicardial region migrating towards underlying myocardium. Scale bars in **f**, 20 μm; **g**, 10 μm. **h**, Three-dimensional Imaris reconstruction of migrating YFP$^+$ cells (green) amidst non-labelled cells (red). ep, epicardium; my, myocardium. **i–l**, YFP$^+$ cells that co-stained for Isl1 (highlighted by white arrowheads in **k**) resided in the epicardium proximal to areas of scarred myocardium 2 days (d2) after myocardial infarction. sc, scar region. Scale bar in **l** (also applies to **i–l**), 50 μm. **m**, Significant increase in *Isl1* expression in primed hearts at days 2, 4 and 7 after myocardial infarction relative to sham-operated controls $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$; MI group 1 and group 2 versus sham; myocardial infarction categories: purple, mild injury; cream, severe injury; $n = 6$ hearts per sham and MI groups. **n**, Significant increases in Isl1$^+$/YFP$^+$ cells at days 2 ($*P \leq 0.05$), 4 ($**P \leq 0.01$) and 7 ($***P \leq 0.001$) after myocardial infarction and Nkx2-5$^+$/YFP$^+$ cells by day 7 ($***P \leq 0.001$) alongside phospho-histone H3$^+$ (P-HH3$^+$) proliferating YFP$^+$ progenitors at day 7 ($***P \leq 0.001$), compared to sham-operated controls. *P* values were calculated by Student's *t*-test (**m**) and paired ANOVA (**n**). Error bars represent mean ± s.e.m. *N* values are numbers of hearts analysed for each group: $N = 3$ (**m**); $N = 4$ (d2 and d4) and $N = 7$ (d7) (**n**).

expanded in number from day 2–14 (Supplementary Fig. 7g–m). A subpopulation, residing in the epicardium, was positive for Isl1 and Nkx2-5 at day 2 after myocardial infarction (Fig. 1i–m and Supplementary Fig. 8a–d) and these cardiac progenitors significantly increased by day 7 after myocardial infarction (Fig. 1n). By day 14, YFP$^+$ cells were located within the border zone and peri-infarct region (Fig. 2a–c), which co-expressed SαA (Fig. 2d, e, h–k) and cTnT (Fig. 2f, g, l) and by virtue of their size, gross morphology and inherent ultrastructure resembled mature cardiomyocytes (Fig. 2d–l). Whereas a relatively small number of Isl1$^+$ progenitors were evident in Tβ4-primed, sham-operated hearts (Fig. 1n), we failed to locate any mature GFP$^+$/YFP$^+$ cardiomyocytes in the absence of injury (sham) or Tβ4 priming (data not shown). The *de novo* cardiomyocytes appeared appropriately integrated with the resident myocardium and with each other, as determined by both N-cadherin (Ncad)$^+$ adherens junctions (Fig. 2f–i) and connexin 43 (Cx43)$^+$ gap junction formation (Fig. 2j, k). Structurally coupled mature cTnT$^+$/YFP$^+$ cardiomyocytes were evident proximal to the scar and within the border zone (Fig. 2l). Consistent with tracking labelled cardiomyocytes from progenitors, proliferating BrdU$^+$ cells at day 4 (Supplementary Fig. 9a–d) were traced to BrdU$^+$/YFP$^+$ cardiomyocytes at day 14 (Supplementary Fig. 9e–h). The total mean percentage of YFP$^+$ progenitors that became cardiomyocytes was 0.59 ± 0.18% (serial sections through $n = 7$ hearts ± s.e.m.). The mean ratio of YFP$^+$ *de novo* cardiomyocytes to YFP$^-$ pre-existing cardiomyoctes residing in the peri-infarct region (Fig. 2b) was 0.066 (1:15; ± 0.0015; six hearts analysed). In addition, from the two-photon imaging we were able to identify numbers of YFP$^+$ cardiomyocytes (192.5 ± 12.1; four hearts analysed) in the underlying myocardium of the left ventricular wall up to a depth of ≤100 μm. Approximately 82% of YFP$^+$ cardiomyocetes were located proximal to the site of injury, residing within either the border zone or the immediate surrounding healthy myocardium (Fig. 2c, l and Supplementary Fig. 10a–c).

To assess functional integration with resident myocardium, we recorded cellular calcium transients $[Ca^{2+}]_i$ between YFP$^+$ and YFP$^-$ cells *in situ*, as previously described[13] (Fig. 2m–s and Supplementary Fig. 10a–f). Two-photon imaging confirmed migration of YFP$^+$ cells from the outer epicardial layer into the underlying myocardium (Supplementary Fig. 7m). At day 14 after myocardial infarction, evoked $[Ca^{2+}]_i$ transients in YFP$^+$ cardiomyocytes were synchronous, with kinetics indistinguishable from those of neighbouring YFP$^-$ cardiomyocytes (Fig. 2q–s). Apparent differences in resolution between resident YFP$^-$ and *de novo* YFP$^+$ cardiomyocyte transients (compare Fig. 2q with r) were observed, reflecting the newly acquired function of the YFP$^+$ population.

To rule out the possibility that we traced resident cardiomyocytes that were labelled by virtue of ectopic activation of the fluorophore from the *Wt1* knock-in alleles, we carried out a series of experiments transplanting FACS-isolated donor GFP$^+$ cells into non-transgenic host hearts. Extensive analyses indicated that the prospective donor cells were progenitors of epicardial origin (Fig. 3). Wt1$^+$ cells were restricted to the epicardium and subepicardial region throughout the heart, as confirmed by co-staining with an antibody against podoplanin, a transmembrane glycoprotein[14] that specifically marked the

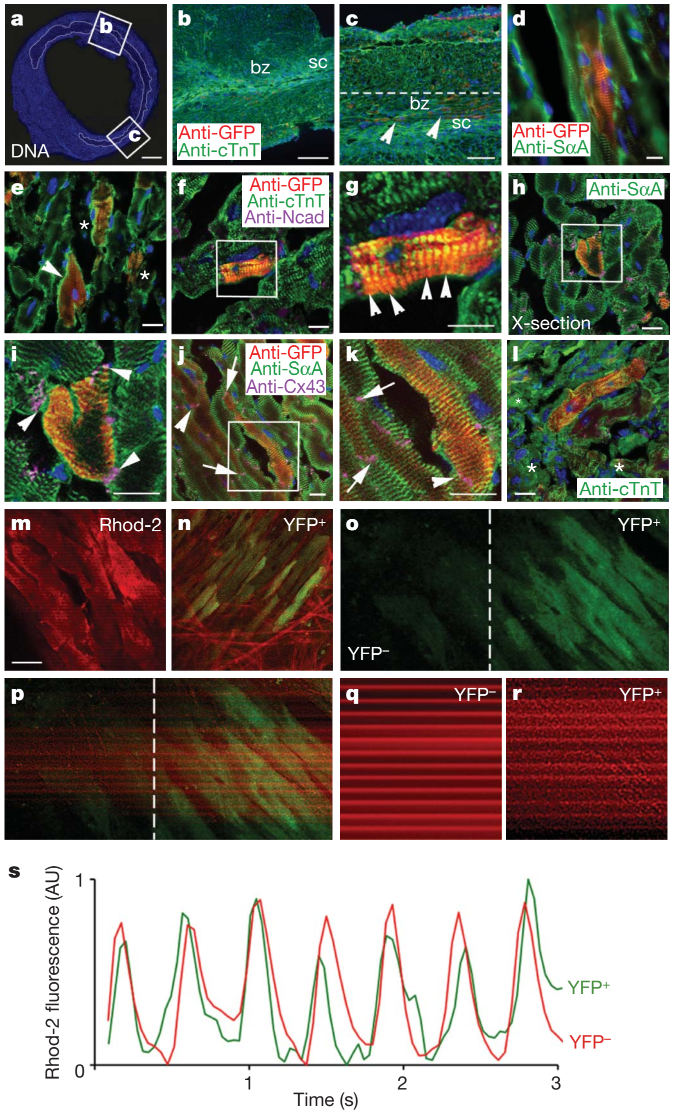


Figure 2 | Activated adult Wt1^{+} progenitors differentiate into structurally coupled cardiomyocytes. a–c, At day 14, YFP^{+} cells that co-expressed cTnT resided in the left ventricular wall within the border zone (white arrowheads in **c**; dashed white line demarcates extent of peri-infarct region). bz, border zone, sc, scar region. **d**, **e**, YFP^{+} cells expressing SαA as determined by epifluorescence (**d**) and confocal microscopy (**e**; white arrowhead highlights mature YFP^{+} cardiomyocyte; white asterisks highlight less mature YFP^{+} cardiomyocytes). **f**, **g**, Mature YFP^{+} cardiomyocytes stained positive for cTnT with sarcomeric banding (white arrowheads). **f–k**, Evidence of structural coupling between YFP^{+} and resident YFP^{-} cardiomyocytes through Ncad^{+} adherens junction (white arrowhead in **i**) and Cx43^{+} gap junction formation (white arrowhead/arrows in **j**, **k**). **l**, YFP^{+}/cTnT^{+} cardiomyocytes were located adjacent to necrotic myocardium within the scar (white asterisks). **m**, **n**, Rhod-2 loading of distal YFP^{-} cardiomyocytes (**m**) was compared against YFP^{+} cardiomyocytes within the peri-infarct region (**n**). **o**, **p**, Calcium transients across clustered YFP^{-} and YFP^{+} cardiomyocytes as evidence of functional coupling. **q**, **r**, Distal YFP^{-} spontaneous calcium transients (**q**) were compared against YFP^{+} transients (**r**). **s**, Representative traces plotted per cardiac cycle (AU, arbitrary units). All scale bars are 20 µm, except for **a**, 100 µm; **b**, 150 µm; and **c**, 150 µm. Scale bar in **m** applies to **m–r**.

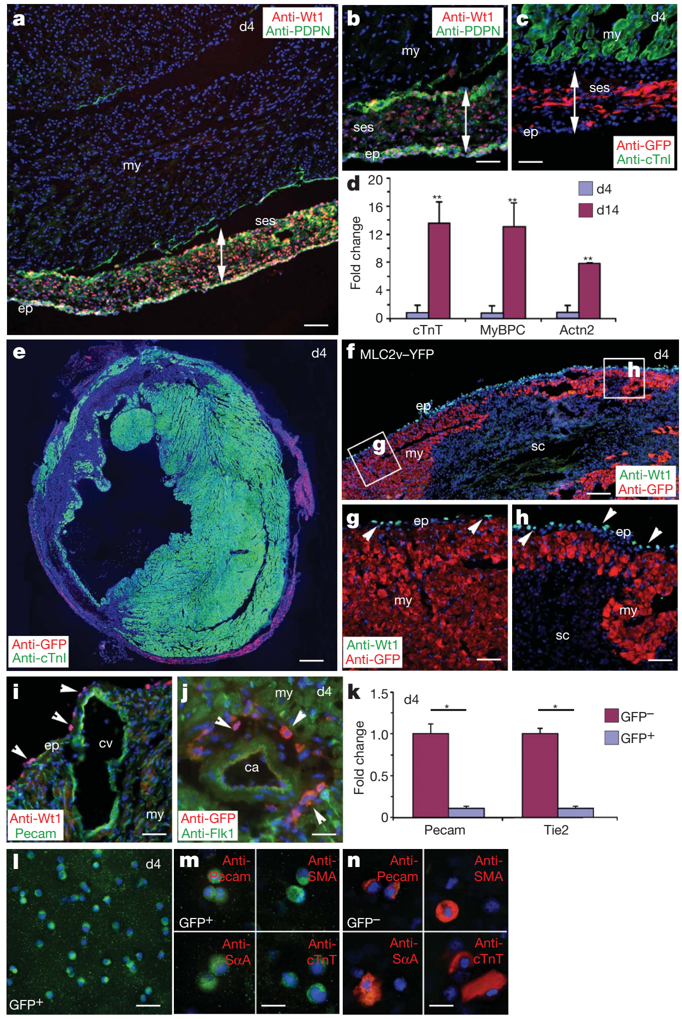


Figure 3 | Prospective donor Wt1^{+}/GFP^{+} cells at day 4 after myocardial infarction seem to be derived from epicardium. a, **b**, Immunostaining for Wt1 and anti-podoplanin (PDPN) revealed that the day 4 donor Wt1^{+} cells were restricted to the epicardium and subepicardial space. ep, epicardium; my, myocardium; sc, scar; ses, subepicardial space. **c**, **e**, Anti-GFP and anti-cTnI co-staining revealed spatial restriction to the equivalent epicardial regions as for Wt1 staining with exclusion from the myocardium. **d**, cTnT, MyBPC and Actn2 were not expressed in the FACS GFP^{+} population at day 4 after myocardial infarction, whereas expression of all three myocardial markers was significantly upregulated in GFP^{+} cells at day 14 consistent with contribution of *de novo* cardiomyocytes. **f–h**, Wt1^{+} cells were restricted to the epicardium and excluded from MLC2v–YFP^{+} ventricular cardiomyocytes in regions of healthy (**g**) and scarred (**h**) myocardium. **i–k**, Expression of Wt1/GFP was excluded from within the Pecam^{+}/Flk1^{+} coronary vasculature (**i**, **j**; cv, coronary vein; ca, coronary artery) and confirmed by a lack of expression of Pecam and Tie2 in the day 4 GFP^{+} FACS population (**k**). **l**, **m**, Cytospin/immunostaining revealed a homogeneous GFP^{+} fraction at day 4, without contamination from GFP^{-} cells (**l**), which lacked vascular (Pecam and SMA) and myocardial (SαA, cTnT) markers (**m**) as compared to the GFP^{-} fraction (**n**). $*P \leq 0.05$, $**P \leq 0.01$, all statistics by Students *t*-test. Error bars represent mean ± s.e.m., N values are numbers of hearts analysed for each group: $N = 6$ (day 4) and $N = 7$ (day 14) (**d**); $N = 6$ (**k**). All scale bars are 50 µm, except **a**, **e**, 500 µm; **f**, 200 µm; **m**, **n**, 20 µm.

epicardial and myocardial boundaries (Fig. 3a, b). Immunostaining for anti-GFP revealed GFP^{+} cells residing in the expanded epicardium but excluded from the myocardium throughout the heart (Fig. 3c, e). Real-time quantitative polymerase chain reaction analyses on FACS-isolated donor cells at day 4 revealed no expression of the canonical cardiomyocyte markers cTnT, MyBPC and Actn2 (Fig. 3d). We also analysed hearts isolated from Tβ4-primed/injured MLC2v$^{Cre/+}$;R26R$^{EYFP/+}$ mice (MLC2v–YFP; Fig. 3f), which lineage-traced ventricular cardiomyocytes as YFP^{+} from early developmental stages[15] to adulthood, and confirmed an absence of Wt1 in both healthy and scarred myocardium respectively (Fig. 3g, h).

In a previous study, Wt1 was shown to be expressed in the coronary vasculature after myocardial infarction in the rat heart[16]. Co-immunostaining for Wt1 or GFP with Pecam and Flk1, respectively, excluded Wt1 expression in both coronary veins and arteries within the donor

hearts and ruled out the possibility of the GFP$^+$ progenitor population arising from existing vasculature (Fig. 3i, j). The FACS-isolated cells at day 4 revealed no expression of the canonical vascular markers Pecam and Tie2 (Fig. 3k), and cytospin with comparative immunostaining of the sorted GFP$^+$ and GFP$^-$ populations (Fig. 3l) revealed that the GFP$^+$ cells were negative for both vascular markers Pecam and SMA and myocardial markers SαA and cTnT (Fig. 3m), whereas the GFP$^-$ population contained cardiovascular cells (Fig. 3n).

After donor cell transplantation (Fig. 4a), GFP$^+$ cells within the epicardial region of the host were restricted to the site of injection after 24 h (Fig. 4b). By day 7, GFP$^+$/Nkx2-5$^+$ donor cells, indicative of myocardial progenitor commitment, were located proximal to the sub-epicardium (Fig. 4c), in conjunction with morphologically immature, cTnT$^+$ cardiomyocyte-like cells (Fig. 4d). Collectively, the presence of these staged donor derivatives suggested progressive differentiation towards a mature cardiomyocyte fate. More definitive donor GFP$^+$ cardiomyocytes with myofibrillar structure and that co-expressed cTnT were observed residing within the host myocardium at day 14 (Fig. 4e, f). We subsequently traced donor GFP$^+$ cardiomyocytes for fluorescent *in situ* hybridization (FISH) with X- and Y-chromosome paints to assess karyotype (Fig. 4a). Single XY GFP$^+$ cardiomyocytes were detected within the XX host, indicating that transdifferentiation had occurred in the absence of cell fusion (predicted XYXX fusion karyotype; Fig. 4g–j). To exclude the possibility of reductive divisions of fusion hybrids accounting for the XY diploid karyotype, reciprocal transplantation experiments injecting female XX GFP$^+$ donor cells into male XY wild-type hosts (Fig. 4k) resulted in XX GFP$^+$/SαA$^+$ cardiomyocytes (Fig. 4l, m) in XY host myocardium (Fig. 4n, o). Quantitative assessment of FISH on GFP$^+$ cardiomyocytes (total $n = 22$) excluded the presence of a host Y chromosome in each case. Finally, we reanalysed non-transplanted YFP$^+$ cells in female hearts (see Fig. 2) and detected examples of both YFP$^+$/cTnT$^+$ (Fig. 4p) and YFP$^+$/SαA$^+$ (Fig. 4r) cardiomyocytes with a single XX karyotype (Fig. 4q, s).

To investigate the outcome of Tβ4 priming on cardiac function and myocardial regeneration, we carried out serial magnetic resonance imaging (MRI) 7, 14 and 28 days after myocardial infarction (Supplementary Fig. 11a–e and Supplementary Table 1). Significant improvement in functional parameters, including ejection fraction and end diastolic/systolic volumes, alongside beneficial changes in infarct/scar volume with increased left ventricular mass over time (Supplementary Table 1) were recorded with Tβ4 treatment, as a surrogate indicator of replenished myocardium (Supplementary Fig. 11f, g).

Collectively, these data indicate that the adult heart can respond to injury with a modest increase in Wt1$^+$ progenitors but without initiating a cardiogenic program. Tβ4 enhances this response, via a precocious and significant reactivation of *Wt1* expression ultimately resulting in cardiomyocyte restitution. Although we cannot unequivocally exclude the possibility that Wt1$^+$ progenitors arise from a nonepicardial source, support for an adult EPDC myocardial contribution comes from the tight regulation of Wt1$^+$ labelling in the epicardium and subepicardial region, and transplantation of donor GFP$^+$ progenitors residing both within and immediately proximal to the epicardium.

Reactivation of *Wt1* by injury and Tβ4 represents a robust means to faithfully tag the progression of adult cardiac progenitors to differentiated myocytes, and provides mechanistic insight into a molecular function of Tβ4 and downstream cellular events. Previously, Tβ4 was shown to upregulate ILK and Akt activity in the heart, enhancing early myocyte survival after ischaemic injury[17]. In addition, Tβ4 can induce the adult epicardium, in the same setting, to contribute coronary endothelial and smooth muscle cells and initiate vascular repair[6,18]. We now propose a further contribution, underpinning the initiation and migration of resident cardiovascular progenitors towards a cardiomyocyte fate. Although in the current study it is not possible to discriminate the relative contribution of each these proposed functions, several effects of Tβ4 seem to be delineated with time post-myocardial

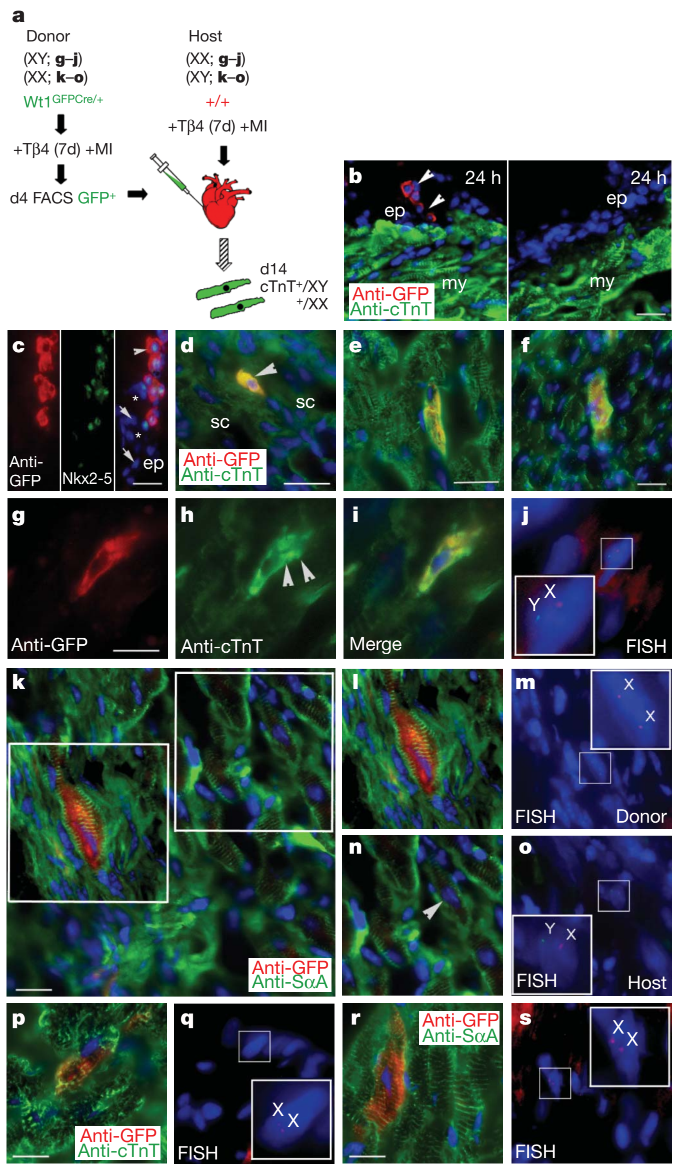


Figure 4 | Transplanted donor Wt1$^+$ progenitors differentiate into cardiomyocytes within host myocardium in the absence of cell fusion. **a**, Schematic of cell transplantation regimen. **b**, After 24 h post-transplantation GFP$^+$ cells within the epicardium and subepicardial region at the injection site were absent in remote regions. ep, epicardium; my, myocardium. **c**, Transplanted GFP$^+$ cells expressed Nkx2-5, indicative of a myocardial progenitor phenotype (white arrowhead highlights a GFP$^+$/Nkx2-5$^+$ progenitor; white arrows highlight GFP$^-$/Nkx2-5$^+$ progenitors and asterisks highlight epicardium cells negative for both GFP and Nkx2-5). **d**–**f**, Donor GFP$^+$ cells with an intermediate differentiated phenotype (highlighted by white arrowhead, **d**), alongside those with evidence of sarcomeric banding that co-expressed cTnT were observed within host myocardium (**e**, **f**). sc, scar. **g**–**j**, GFP$^+$/cTnT$^+$ cardiomyocytes (**g**–**i**) with sarcomeric banding (highlighted by white arrowheads in **h**) were traced for FISH analyses to reveal a single XY karyotype (**j**). X, X chromosome; Y, Y chromosome. **k**–**o**, Reciprocal transplantation (XX into XY) followed by confocal microscopy (**k**) revealed GFP$^+$ cardiomyocytes (**l**) that had the donor XX karyotype (**m**) relative to the XY karyotype of host GFP$^-$ cardiomyocytes (highlighted by white arrowhead; **n**, **o**). **p**–**s**, In female Wt1$^{CreERT2/+}$;R26R$^{EYFP/+}$ mice, previously tracked YFP$^+$ cTnT$^+$ (**p**) and SαA$^+$ cardiomyocytes (**r** and examples shown in Fig. 3), had an XX karyotype (white box insets) supporting transdifferentiation in the absence of cell fusion with resident cardiomyocytes (**q**, **s**). Scale bars: **c**, 25 μm; **d**–**f**, 20 μm; **g** (applies to **g**–**j**), **k** (applies to **k**–**o**), **p** (applies to **p**–**q**), **r** (applies to **r**–**s**), 10 μm.

infarction. The pro-survival activity of Tβ4 is an early injury response to maintain the status quo of surviving myocardium, whereas the neovascularization and *de novo* cardiogenesis are longer-term regenerative functions potentially acting through the common target of adult EPDCs.

The identification of a bona fide source of myocardial progenitors is a significant step towards resident-cell-based therapy for acute myocardial infarction in human patients. The induced differentiation of the progenitor pool described into cardiomyocytes by Tβ4 is at present an inefficient process relative to the activated progenitor population as a whole. Consequently, the search is on via chemical and genetic screens to identify efficacious small molecules and other trophic factors to underpin optimal progenitor activation and replacement of destroyed myocardium.

METHODS SUMMARY

MLC2v–YFP and inducible $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ mice were generated by crossing Rosa26R–EYFP with $MLC2v^{Cre/+}$ (ref. 15) and $Wt1^{CreERT2/+}$ (ref. 4) mice. *Wt1* expression in adult epicardium was achieved through daily intraperitoneal Tβ4 injections (RegeneRX; 12 mg kg^{-1}, 7 days; Fig. 1a).$Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ mice received 2 mg tamoxifen. EPDC cultures from Tβ4-primed adult $Wt1^{GFPCre/+}$ mice were differentiated for 14 days in IMDM plus 20% FBS and fixed in 4% paraformaldehyde. Permanent left anterior descending artery (LAD) ligation was performed in adult $Wt1^{GFPCre/+}$ ($n = 85$, sham $n = 6$), $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ ($n = 31$, sham $n = 6$), MLC2v–YFP ($n = 5$), wild-type transplantation hosts ($n = 5$ females; $n = 6$ males) and wild-type C57BL/6J mice ($n = 21$). Tβ4 or vehicle was given on alternate days post-myocardial infarction. Two-hundred microlitres of BrdU labelling reagent (Invitrogen) was injected at days 0, 2 and 4. Hearts were digested with 0.1% collagenase II, GFP^{+}/$EYFP^{+}$ cells quantified (Beckman Coulter MoFlo XDP) and c-Kit and Sca-1 expression determined (Beckman Coulter CyAn ADP). Cytospins were collected using a Shandon Cytospin 3 centrifuge. Immunofluorescence analysis was performed on EPDC cultures, cytospun FACS-sorted cells and post-myocardial-infarction heart cryosections. Total RNA was isolated from myocardial infarction heart apex and from FACS-sorted GFP^{+} cells and reverse transcribed for real-time qRT–PCR using SYBR Green on an ABI 7900 with *Hprt* as endogenous control. Fold-change was determined by the $2^{-\Delta\Delta CT}$ method and presented relative to levels in sham hearts. Multiphoton imaging was performed on Langendorff perfused hearts loaded with rhod-2/AM as described[19]. Spontaneous calcium transients, visualized without electrical pacing, were calculated using ImageJ. MRI analysis was performed on $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ and wild-type C57BL/6J mice, treated with Tβ4 or vehicle, 7, 14 and 28 days post-myocardial infarction, to determine temporal changes in infarct size and cardiac function. For transplantation experiments 3–6 × 10^4 FACS-sorted GFP^{+} cells were injected under the epicardium of sex mismatched, LAD-ligated, Tβ4-treated mice. Hearts harvested at 24 h, 7 days and 14 days were processed for immunofluorescence and FISH.

All animal experiments were carried out according to UK Home Office project licence PPL 70/6691 compliant with the UK Animals (Scientific Procedures) Act 1986 and approved by the University College London Biological Services Ethical Review Process.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 March 2010; accepted 13 May 2011.
Published online 8 June 2011.

1. Willems, E., Bushway, P. J. & Mercola, M. Natural and synthetic regulators of embryonic stem cell cardiogenesis. *Pediatr. Cardiol.* **30,** 635–642 (2009).
2. Smart, N. *et al.* Thymosin β4 induces adult epicardial progenitor mobilization and neovascularization. *Nature* **445,** 177–182 (2007).
3. Cai, C. L. *et al.* A myocardial lineage derives from *Tbx18* epicardial cells. *Nature* **454,** 104–108 (2008).
4. Zhou, B. *et al.* Epicardial progenitors contribute to the cardiomyocyte lineage in the developing heart. *Nature* **454,** 109–113 (2008).
5. Christoffels, V. M. *et al.* Tbx18 and the fate of epicardial progenitors. *Nature* **458,** E8–E9 (2009).
6. Bock-Marquette, I. *et al.* Thymosin β4 mediated PKC activation is essential to initiate the embryonic coronary developmental program and epicardial progenitor cell activation in adult mice *in vivo*. *J. Mol. Cell. Cardiol.* **46,** 728–738 (2009).
7. Smart, N. & Riley, P. R. Derivation of epicardium-derived progenitor cells (EPDCs) from adult epicardium. *Curr. Protoc. Stem Cell Biol.* Unit 2C.2 (2009).
8. Laugwitz, K. L. *et al.* Postnatal $isl1^{+}$ cardioblasts enter fully differentiated cardiomyocyte lineages. *Nature* **433,** 647–653 (2005).
9. Moretti, A. *et al.* Multipotent embryonic $Isl1^{+}$ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell* **127,** 1151–1165 (2006).
10. Wu, S. M. *et al.* Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart. *Cell* **127,** 1137–1150 (2006).
11. Prall, O. W. *et al.* An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. *Cell* **128,** 947–959 (2007).
12. Limana, F. *et al.* Myocardial infarction induces embryonic reprogramming of epicardial $c\text{-}kit^{+}$ cells: role of the pericardial fluid. *J. Mol. Cell. Cardiol.* **48,** 609–618 (2010).
13. Rubart, M. *et al.* Two-photon molecular excitation imaging of Ca^{2+} transients in Langendorff-perfused mouse hearts. *Am. J. Physiol. Cell Physiol.* **284,** C1654–C1668 (2003).
14. Mahtab, E. A. *et al.* Cardiac malformations and myocardial abnormalities in podoplanin knockout mouse embryos: correlation with abnormal epicardial development. *Dev. Dyn.* **237,** 847–857 (2008).
15. Chen, J., Kubalak, S. W. & Chien, K. R. Ventricular muscle-restricted targeting of the RXRα gene reveals a non-cell-autonomous requirement in cardiac chamber morphogenesis. *Development* **125,** 1943–1949 (1998).
16. Wagner, K. D. *et al.* The Wilms' tumor suppressor *Wt1* is expressed in the coronary vasculature after myocardial infarction. *FASEB J.* **16,** 1117–1119 (2002).
17. Bock-Marquette, I. *et al.* Thymosin β4 activates integrin-linked kinase and promotes cardiac cell migration, survival and cardiac repair. *Nature* **432,** 466–472 (2004).
18. Smart, N. *et al.* Thymosin β4 facilitates epicardial neovascularization of the injured adult heart. *Ann. N. Y. Acad. Sci.* **1194,** 97–104 (2010).
19. Rubart, M. *et al.* Two-photon molecular excitation imaging of Ca^{2+} transients in Langendorff-perfused mouse hearts. *Am. J. Physiol. Cell Physiol.* **284,** C1654–C1668 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by the British Heart Foundation. We are grateful to F. Costantini and S. Srinivas for providing the R26R–EYFP mouse strain, to B. Vernay for assistance with confocal microscopy and A. Eddaoudi, P. Chana and A. Angheluta for assistance in flow cytometry. We thank A. Taylor and V. Muthurangu for functional interpretation of MRI data and RegeneRx Biopharmaceuticals for provision of clinical grade Tβ4.

Author Contributions N.S. carried out the *in vivo* histological assessments of cardiomyocytes and FISH experiments. S.B. carried out the explant and FACS studies and jointly with K.N.D. established the myocardial infarction model and the cell transplantation. J.M.V. carried out the qRT–PCR analyses and assisted with cell transplantation. B.Z. generated the $Wt1^{GFPCre}$ and $Wt1^{CreERT2}$ mice. S.D. and D.Y. performed the two-photon microscopy and Ca^{2+} transient recordings. J.R., A.N.P. and M.F.L. carried out the MRI functional analyses. W.T.P. provided the $Wt1^{GFPCre}$ and $Wt1^{CreERT2}$ mice. P.R.R. established the hypotheses and experimental design, co-analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.R.R. (p.riley@ich.ucl.ac.uk).

METHODS

Generation of epicardial trace mice. $Wt1^{GFPCre/+}$ mice have been previously described[4]. Inducible $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ mice were generated by crossing the $Wt1^{CreERT2/+}$ strain with $Rosa26R^{EYFP/+}$ reporter mice and genotyping as previously described. Adult mice were primed with intraperitoneal (i.p.) injection of Tβ4 or vehicle (PBS) into either $Wt1^{GFPCre/+}$ or $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ (plus tamoxifen) strains. Primed mice were subsequently given a myocardial infarction by coronary artery ligation ($n = 85$, $Wt1^{GFPCre/+}$ and $n = 31$, $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$). Half of each myocardial infarction group were pre-treated with Tβ4 and half were vehicle treated ($n = 6$ sham-operated controls were included per experiment). Hearts were subsequently assessed using a combination of FACS, cytospin, immunofluorescence and real-time qRT–PCR analyses for GFP/YFP expression and myocardial markers after 2, 4, 7 and 14 days (see schematic in Fig. 1a). With respect to monitoring progenitor-derived cardiomyocytes the focus was on the inducible $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ model to ensure specific temporal labelling of YFP^+ derivatives and rule out ectopic activation of the *Wt1* gene at the targeted allele. Mice that were PBS treated with tamoxifen and Tβ4 primed in the absence of tamoxifen were used as controls and importantly we never observed YFP^+ cardiomyocyte-like cells in these hearts.

Generation of $MLC2v^{Cre/+}$; $R26R^{EYFP/+}$ mice. $MLC2v^{Cre/+}$ mice have been described previously[15] and were crossed with the $R26R^{EYFP/+}$ strain to generate MLC2v–YFP mice for myocardial infarction and to exclude Wt1 upregulation in YFP^+ cardiomyocytes.

Adult epicardial explant cultures. Adult EPDCs were prepared, as previously described[7], from 8–10-week-old $Wt1^{GFPCre/+}$ mice that had received daily i.p. injections of Tβ4 (RegeneRX, $12\,mg\,kg^{-1}$ in PBS) or vehicle (PBS) for 7 days. Cells were allowed to differentiate for up to 14 days in Iscove's modified Dulbecco's medium (IMDM) containing 20% FBS, before fixing in 4% paraformaldehyde for immunostaining analysis. At the outset GFP^+ cells were recorded emerging from Tβ4-treated explants (66.8 ± 4.5; mean percentage of GFP^+ cells relative to total number of cells in outgrowth ± s.e.m.; $n = 12$ explants). Outgrowing cells, up to 6 days in culture, were documented as immature and phenotypically similar to $Nkx2\text{-}5^+$ progenitors previously isolated from embryonic hearts[10] (Supplementary Fig. 2a–d). Throughout the *ex vivo* studies, PBS (vehicle)-treated explants were used as controls and revealed limited outgrowth or emergence of a significantly reduced number of GFP^+ progenitor-like cells. The percentage incidence of GFP^+ progenitors in the PBS treated/control explant cultures was recorded as $14.3 \pm 1.9\%$ (mean ± s.e.m.; $n = 12$), significantly lower ($P \leq 0.001$) than those primed by Tβ4. Importantly, vehicle-treated cells failed to adopt a myocardial fate and only isolated fibroblast-derivatives were observed in control cultures (not shown).

Despite evidence of sarcomeric marker expression (cTnT, SαA and MyBPC) after Tβ4 treatment we did not observe beating in culture of the more differentiated cardiomyocyte-like cells documented. Because spontaneous contraction *in vitro* is confined to immature fetal or neonatal cardiomyocytes[20], this may reflect a more mature adult myocyte phenotype.

Tβ4 administration. The injection regimen of Tβ4 for priming $Wt1^{GFPCre/+}$ and $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ mice, including induction with tamoxifen is outlined in Fig. 1a. In separate experiments, $MLC2v^{Cre/+}$;$R26R^{EYFP/+}$ and wild-type C57BL/6J mice were subject to the same injection regimen for epicardial priming. Adult mice received i.p. injection of Tβ4 ($12\,mg\,kg^{-1}$) or vehicle (PBS) daily for 7 days. On the eighth day $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ mice were injected with tamoxifen (2 mg suspended in peanut oil; i.p.) to induce CreERT2/Cre expression. Further injections of Tβ4/vehicle were given on day 9 and tamoxifen on day 10. Myocardial infarction was performed 3 days after this regimen.

Myocardial infarction. Mice were housed and maintained in a controlled environment. All surgical and pharmacological procedures were performed in accordance with the Animals (Scientific Procedures) Act 1986, (Home Office, UK). For all experiments adult male $Wt1^{GFPCre/+}$ ($n = 85$), $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ ($n = 31$), $MLC2v^{Cre/+}R26R^{EYFP/-}$ mice ($n = 5$) and wild-type C57BL/6J mice ($n = 21$), (25–30 g) were primed with Tβ4 or vehicle, as described earlier, before surgical procedures. Myocardial infarction was induced in isoflurane-anaesthetized mice by permanent ligation of the left anterior descending artery (LAD). Sham controls (suture passed under the LAD but not ligated) were performed in $Wt1^{GFPCre/+}$ mice ($n = 6$) and $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ mice ($n = 6$). On recovery, animals received i.p. injection of Tβ4 ($12\,mg\,kg^{-1}$) or vehicle (PBS). Further injections were given every second day. For pulse chase BrdU experiments, $Wt1^{GFPCre/+}$ and $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ mice were injected i.p. with 200 μl of BrdU labelling reagent (Zymed, Invitrogen) on recovery and at 2 and 4 days after LAD ligation. Hearts were harvested at 2, 4, 7 and 14 days after ligation and bisected transversely midway through the scar: the apex was snap frozen for RNA isolation and subsequent real-time qRT–PCR studies while the remaining tissue was fixed in 4% PFA for cryosectioning and immunostaining analyses.

GFP^+/YFP^+ cell isolation and characterization. Hearts from $Wt1^{GFPCre/+}$ and $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ mice treated with Tβ4 or vehicle were harvested 7 days after ligation and processed by enzymatic digestion using a 0.1% collagenase II-PBS solution (Worthington Biochemicals) to achieve a single-cell suspension. GFP^+ or $EYFP^+$ cells were isolated from the total cardiac cell population using a Beckman Coulter MoFlo XDP cell sorter with a 488 nm laser beam to excite YFP/GFP (collected in the 530/40 nm channel) and a 355 nm laser beam used to excite DAPI (collected in the 450/50 nm channel).

To characterize GFP^+ cells, hearts from $Wt1^{GFPCre/+}$ mice treated with Tβ4 were harvested 4 days after ligation and processed, as described earlier, to obtain a single-cell suspension. As a control, hearts from uninjured $Wt1^{GFPCre/+}$ mice treated either with Tβ4 or with vehicle were also analysed after 7 days. Cells were incubated with the following primary and secondary antibodies: c-Kit (goat IgG, R&D System), Sca-1 (rat IgG2a, BD Pharmigen), Alexa Fluor 647 anti-goat or Alexa Fluor 647 anti-rat (Invitrogen) and analysed using a Beckman Coulter CyAn ADP analyser equipped with a 488 nm laser and 633 nm red diode and run by Summit Software. Data were analysed using the FlowJo Software.

Cell cytospins were collected using a Shandon Cytospin 3 centrifuge. Cytospun cells were then processed for immunostaining for epicardial and cardiovascular markers as described earlier.

YFP^+ cardiomyocytes were assessed by cell counts through serial sections. The mean percentage of YFP^+ progenitors that became cardiomyocytes was estimated across $n = 7$ hearts ± s.e.m. The incidence of YFP^+ *de novo* cardiomyocytes relative to pre-existing myocardium was expressed as the mean ratio of YFP^+ cells divided by YFP^- cells across serial sections per heart ± s.e.m. ($n = 6$ hearts analysed). Cell counts were also assessed after two-photon imaging in Langendorff-perfused hearts and expressed as the mean number of YFP^+ cardiomyocytes per heart ± s.e.m. ($n = 4$ hearts analysed).

Immunodetection methods. Immunofluorescence was performed on adult epicardial explant cultures, on cytospun FACS-sorted cells and on cryosections of post-myocardial infarction hearts using standard protocols with the following antibodies: GFP (Clontech and Abcam, which also detect EYFP), SαA (Sigma), cardiac MyBPC (a gift from E. Ehler and M. Gautel), cTnT, cTNI, BrdU, Isl1 and Wt1 (all Abcam), Nkx2-5 (Santa Cruz), phospho-histone H3 (Upstate), podoplanin (Fitzgerald Industries) and Ki67 (Dako). To rule out the possibility of autofluorescence accounting for the detection of either GFP or YFP protein expression, sections through the left ventricle were stained with a polyclonal anti-GFP antibody (which detects both fluorescent proteins). The specificity of the anti-GFP antibody was ascertained by immunofluorescence on non-primed, intact hearts, which detected neither labelled cells in the epicardial region, nor their derivatives (no signal; not shown).

To detect BrdU-positive nuclei, sections were treated with 2 N HCl for 30 min at room temperature (22 °C) to denature the DNA, and neutralized in 0.1 M sodium borate pH 8.5 for 12 min before incubation with the anti-BrdU antibody. Owing to the destruction of cellular antigens resulting from acid treatment, these steps were performed after the incubation with antibodies to GFP and SαA. Images were acquired using either a Zeiss AxioImager with ApoTome or a Zeiss LSM 710 confocal microscope equipped with argon and helium neon lasers using ×20, ×40 and ×63/1.4 (oil immersion) objectives.

RNA *in situ* hybridization. RNA *in situ* hybridization on adult heart cryosections was performed as previously described[21], using a digoxigenin-labelled antisense riboprobe specific for *Wt1* (ref. 4), alongside a sense control.

RNA isolation and gene expression profiling. Total RNA was isolated from the apex of collected hearts using the Trizol reagent (Invitrogen), according to the manufacturer's instructions and reverse-transcribed using Superscript III RT (Invitrogen). Real-time qRT–PCR analysis was performed on an ABI 7900 Sequence Detector (Applied Biosystems) using SYBR Green (QuantitectTM SYBR Green PCR Kit, Qiagen). Data were normalized to Hprt expression (endogenous control). Fold-changes in gene expression were determined by the $2^{-\Delta\Delta CT}$ method[22] and are presented relative to levels in non-myocardial infarction (sham) hearts. Complementary DNA PCR primer sequences were obtained from Primer Bank (http://pga.mgh.harvard.edu/primerbank/) and details are available on request.

To characterize GFP^+ progenitors, total RNA was obtained from FACS-sorted GFP^+ cells isolated after enzymatic digestion of the hearts of the Tβ4-treated $Wt1^{GFPCre/+}$ mice at 4 and 14 days following LAD ligation using a Beckman Coulter MoFlo XDP cell sorter. Total RNA was isolated using the RNeasy Micro Kit (Qiagen), according to the manufacturer's instructions, and processed as above.

Multiphoton imaging. Multiphoton imaging was performed in Langendorff-perfused hearts loaded with Rhod- 2/AM (Invitrogen) as described previously[19] except that 50 μM blebbistatin was used to inhibit myosin crossbridge cycling and prevent movement. Spontaneous calcium transients were visualized without

electrical pacing. Images were recorded using a C-apochromat $\times$40/1.2 NA water-immersion objective on a Zeiss LSM NLO axiovert microscope coupled to a tunable Chameleon laser (Coherent) and external (nondescanned) detectors. EYFP/EGFP and Rhod-2/AM were excited simultaneously at 990 nm, which is long enough to avoid myofibril autofluorescence while strongly exciting the fluorochromes. Emitted light was collected using a bandpass 500–550 nm and 575–640 nm. Calcium transients were calculated by averaging along a line 40 pixels wide using ImageJ (http://rsbweb.nih.gov/ij) software. The ImageJ fourier filter was used to remove noise of less than 3 pixels and three-dimensional images were constructed using Imaris (Bitplane).

MRI analysis. $Wt1^{CreERT2/+}$;$R26R^{EYFP/+}$ and wild-type C57BL/6J mice treated with Tβ4 or vehicle, were subjected to MRI assessment at 7 days after LAD ligation. Where infarct size was within the range of 15–40%, follow-up MRI analysis was performed on the same mice at 14 and 28 days post-myocardial infarction, to determine temporal changes in infarct size and cardiac function. Mice were anaesthetized with isoflurane (4%), placed onto an animal cradle and maintained at $37 \pm 0.5\,^{\circ}$C with oxygen and anaesthetics (1–2% isoflurane), supplied via a nose cone ($1\,\mathrm{l\,min^{-1}}$). Cardiorespiratory monitoring and gating were performed using an MR-compatible system (SA Instruments) with needle electrodes inserted into the front limbs and a respiratory pillow placed on the chest. Imaging was performed using a 9.4T VNMRS horizontal bore scanner (Varian) with a shielded gradient system ($1{,}000\,\mathrm{mT\,m^{-1}}$) using a 39 mm diameter volume coil (Rapid Biomedical GmbH). An electrocardiogram and respiratory gated spoiled gradient echo sequence was used to acquire cine cardiac images with the following parameters for standard cine acquisitions: Time to echo (TE), 1.18 ms; time to repetition (TR), 4.5 ms; flip angle, 20°; slice thickness, 1 mm; no slice separation, field of view (FOV), $25.6 \times 25.6\,\mathrm{mm^2}$; matrix size, 128×128; number of signal averages (NSA), 2. Twenty cine-frames were recorded to cover the cardiac cycle. Infarct size was assessed using late gadolinium enhancement (LGE), as previously described[23]. Briefly, $0.6\,\mathrm{mmol\,kg^{-1}}$ Gd-DTPA was administered i.p. followed by a Look-Locker acquisition with multiple time to inversion (TI) to determine the optimum TI. This was followed by a multi-slice inversion recovery (IR) acquisition with flip angle (FA) = 90° using the following imaging parameters: TE, 1.58 ms; TR, ~500–600 ms; FA, 90°; slice thickness, 0.5 mm; 0.5 mm slice gap; 7–8 slices; FOV, $25.6 \times 25.6\,\mathrm{mm^2}$; matrix size, 192×192; NSA, 2. A second stack of short-axis images offset by 0.5 mm was acquired to generate a continuous data set.

MR image analysis. Randomized and anonymized images were analysed using the cardiac analysis software Segment (http://segment.heiberg.se)[24]. To estimate the infarct size, endocardial and epicardial borders were segmented on LGE images automatically with manual adjustments followed by automatic delineation of infarct tissue using a built-in fraction of segment. Manual corrections were performed where necessary. Infarct size, expressed as percentage of left ventricular mass, was calculated as infarct volume/left ventricular volume (from cine data). Results are shown as mean ± s.e.m. Comparisons between groups were performed using a repeated measures one-way ANOVA. All statistical analysis was performed using R software[25] version 2.8.1.

Cell transplantation. Adult male ($n = 14$) and female ($n = 14$) $Wt1^{GFPCre/+}$ mice were primed with Tβ4 and myocardial infarction was induced as described earlier. Pooled GFP^{+} progenitors were isolated (from eight surviving donors) using a Beckman Coulter MoFlo XDP cell sorter after enzymatic digestion of the hearts of Tβ4-treated $WT1^{GFPCre/+}$ mice 4 days after ligation. Female ($n = 3$) and male ($n = 3$) non-transgenic mice were treated with Tβ4 before surgery and myocardial infarction was induced as described above. $3\text{–}6 \times 10^4$ FACS-sorted male GFP^{+} cells, resuspended in 10 µl of DMEM, were injected into the subepicardial space of the female host hearts ($n = 3$) immediately after LAD ligation. On recovery, animals received i.p. injection of Tβ4 ($12\,\mathrm{mg\,kg^{-1}}$). Further injections were given every second day and hearts were harvested 14 days post-myocardial infarction and processed for immunofluorescence analysis as described earlier. In the second set of experiments, 5×10^4 FACS-sorted female GFP^{+} cells, resuspended in 10 µl of DMEM, were injected into the subepicardial space of the male host hearts ($n = 3$) immediately after LAD ligation. On recovery, animals received i.p. injection of Tβ4 ($12\,\mathrm{mg\,kg^{-1}}$). Further injections were given every second day and hearts were harvested at 24 h, 7 days and 14 days post-myocardial infarction and processed for immunofluorescence analysis.

FISH. After immunofluorescence analysis of EPDC-derived cardiomyocytes in adult heart cryosections, as described earlier, images were acquired before FISH. FISH was performed using mouse StarFISH probes (Cambio), essentially according to the manufacturer's instructions, with the following modifications: muscle was digested in 0.025% pepsin for 40 min at 37 °C. After dehydration, tissue sections were denatured by immersion in 70% formamide at 72 °C for 5 min. Cy3-conjugated X- and FITC-conjugated Y-chromosome paints were mounted, sealed under coverslips and denatured by incubation at 60 °C for 10 min before hybridization in a humid chamber overnight at 37 °C. Coverslips were removed and sections were either washed, according to the manufacturer's instructions, or, where Y-chromosome probes were used, a FITC amplification kit (Cambio) was used, according to the manufacturer's instructions.

20. Ieda, M. *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142,** 375–386 (2010).
21. Moorman, A. F. *et al.* Sensitive nonradioactive detection of mRNA in tissue sections: novel application of the whole-mount *in situ* hybridization protocol. *J. Histochem. Cytochem.* **49,** 1–8 (2001).
22. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25,** 402–408 (2001).
23. Price, A. *et al.* Late gadolinium enhanced MRI in small animal models of myocardial infarction. *J. Cardiovasc. Magn. Reson.* **12** (Suppl. 1), P98 (2010).
24. Heiberg, E. *et al.* Time resolved three-dimensional automated segmentation of the left ventricle. *Comput. Cardiol.* **32,** 599–602 (2005).
25. Ihaka, R. & Gentleman, R. R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* **5,** 299–314 (1996).

doi:10.1038/nature10201

A low mass for Mars from Jupiter's early gas-driven migration

Kevin J. Walsh[1,2], Alessandro Morbidelli[1], Sean N. Raymond[3,4], David P. O'Brien[5] & Avi M. Mandell[6]

Jupiter and Saturn formed in a few million years (ref. 1) from a gas-dominated protoplanetary disk, and were susceptible to gas-driven migration of their orbits on timescales of only ~100,000 years (ref. 2). Hydrodynamic simulations show that these giant planets can undergo a two-stage, inward-then-outward, migration[3–5]. The terrestrial planets finished accreting much later[6], and their characteristics, including Mars' small mass, are best reproduced by starting from a planetesimal disk with an outer edge at about one astronomical unit from the Sun[7,8] (1 AU is the Earth–Sun distance). Here we report simulations of the early Solar System that show how the inward migration of Jupiter to 1.5 AU, and its subsequent outward migration, lead to a planetesimal disk truncated at 1 AU; the terrestrial planets then form from this disk over the next 30–50 million years, with an Earth/Mars mass ratio consistent with observations. Scattering by Jupiter initially empties but then repopulates the asteroid belt, with inner-belt bodies originating between 1 and 3 AU and outer-belt bodies originating between and beyond the giant planets. This explains the significant compositional differences across the asteroid belt. The key aspect missing from previous models of terrestrial planet formation is the substantial radial migration of the giant planets, which suggests that their behaviour is more similar to that inferred for extrasolar planets than previously thought.

Hydrodynamic simulations show that isolated giant planets embedded in gaseous protoplanetary disks carve annular gaps and migrate inward[9]. Saturn migrates faster than Jupiter; if Saturn is caught in the 2:3 mean motion resonance with Jupiter (conditions for this to happen are given in Supplementary Information section 3), where their orbital period ratio is 3/2, generally the two planets start to migrate outward until the disappearance of the disk[3–5,10]. Jupiter could have migrated inward only before Saturn approached its final mass and was captured in resonance. The extents of the inward and outward migrations are unknown a priori owing to uncertainties in disk properties and in relative timescales for the growth of Jupiter and Saturn. Thus we search for constraints on where Jupiter's migration may have reversed (or 'tacked', using a sailing analogy).

The terrestrial planets are best reproduced when the disk of planetesimals from which they form is truncated, with an outer edge at 1 AU (refs 7, 8). These conditions are created naturally if Jupiter tacked at ~1.5 AU. However, before concluding that Jupiter tacked at this distance, a major question needs to be addressed: can the asteroid belt, between 2 and 3.2 AU, survive the passage of Jupiter?

Volatile-poor asteroids (mostly S types) are predominant in the inner asteroid belt, while volatile-rich asteroids (mostly C types) are predominant in the outer belt. These two main classes of asteroids have partially overlapping semimajor axis distributions[11,12], though C types outnumber S types beyond ~2.8 AU. We ran a suite of dynamical simulations to investigate whether this giant planet migration scheme is consistent with the existence and structure of the asteroid belt. Because of the many unknowns in giant planet growth and early dynamical evolution, we present a simple scenario that reflects one plausible history for the giant planets (Fig. 1). We provide an exploration of parameter space (see Supplementary Information) that embraces a large range of possibilities and demonstrates the robustness of the results. In all simulations, we maintain the fundamental assumption that Jupiter tacked at 1.5 AU.

Figure 2 shows how the migration of the giant planets affects the small bodies. The disk interior to Jupiter has a mass 3.7 times that of the Earth ($3.7M_\oplus$), equally distributed between planetary embryos (large)

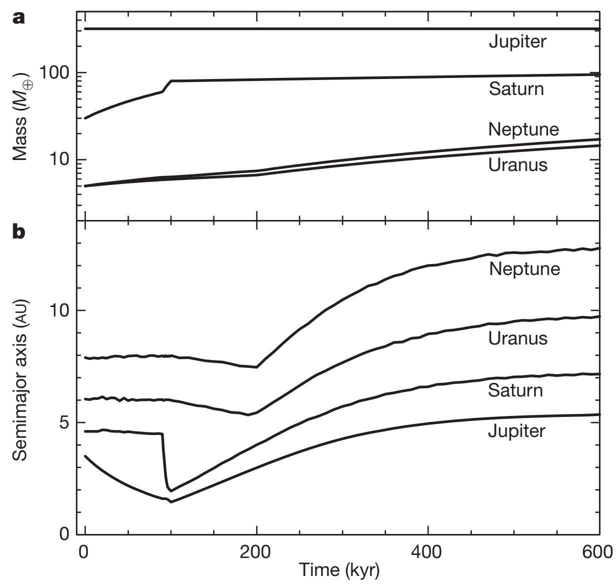


Figure 1 | The radial migration and mass growth imposed on the giant planets in the reference simulation. **a**, Mass growth; **b**, semimajor axis. A fully-formed Jupiter starts at 3.5 AU, a location expected to be highly favourable for giant planet formation owing to the presence of the so-called snow line[21]. Saturn's 30 $M_\oplus$ core is initially at ~4.5 AU and grows to 60 $M_\oplus$ as Jupiter migrates inward, over 10^5 years. Inward type-I migration of planetary cores is inhibited in disks with a realistic cooling timescale[23–26]; thus Saturn's core remains at 4.5 AU during this phase. Similarly, the cores of Uranus and Neptune begin at ~6 and 8 AU and grow from 5 $M_\oplus$, without migrating. Once Saturn reaches 60 $M_\oplus$ its inward migration begins[25], and is much faster than that of the fully grown Jupiter[27]. Thus, on catching Jupiter, Saturn is trapped in the 2:3 resonance[3]. Here this happens when Jupiter is at 1.5 AU. The direction of migration is then reversed, and the giant planets migrate outward together. In passing, they capture Uranus and Neptune in resonance and these planets are then pushed outwards as well. Saturn, Uranus and Neptune reach their full mass at the end of the migration when Jupiter reaches 5.4 AU. The migration rate decreases exponentially as the gas disk dissipates. The final orbital configuration of the giant planets is consistent with their current orbital configuration when their later dynamical evolution is considered[28,29] (see Supplementary Information section 3 for extended discussion).

[1]Université de Nice – Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 4, France. [2]Department of Space Studies, Southwest Research Institute, 1050 Walnut Street, Suite 300, Boulder, Colorado 80302, USA. [3]Université de Bordeaux, Observatoire Aquitain des Sciences de l'Univers, 2 Rue de l'Observatoire, BP 89, F-33270 Floirac Cedex, France. [4]CNRS, UMR 5804, Laboratoire d'Astrophysique de Bordeaux, 2 Rue de l'Observatoire, BP 89, F-33270 Floirac Cedex, France. [5]Planetary Science Institute, 1700 East Fort Lowell, Suite 106, Tucson, Arizona 85719, USA. [6]NASA Goddard Space Flight Center, Code 693, Greenbelt, Maryland 20771, USA.

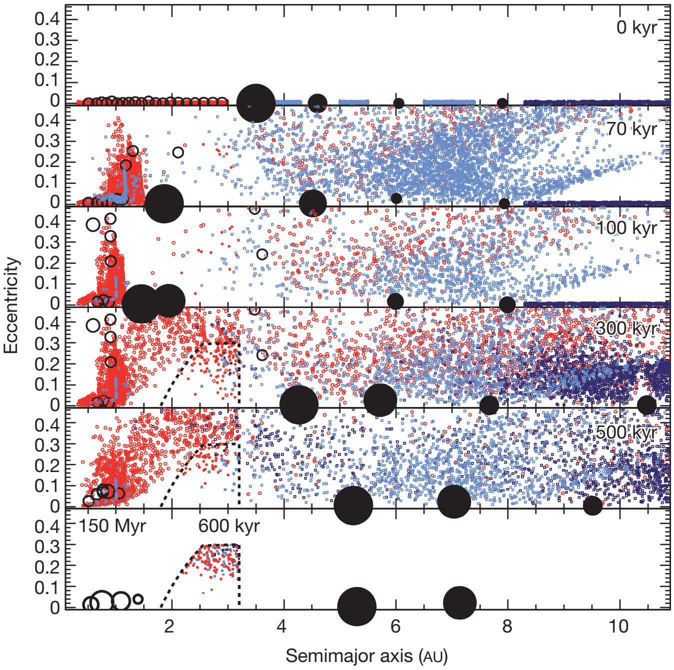


Figure 2 | The evolution of the small-body populations during the growth and migration of the giant planets, as described in Fig. 1. Jupiter, Saturn, Uranus and Neptune are represented by large black filled circles with evident inward-then-outward migration, and evident growth of Saturn, Uranus and Neptune. S-type planetesimals are represented by red dots, initially located between 0.3 and 3.0 AU. Planetary embryos are represented by large open circles scaled by $M^{1/3}$ (but not in scale relative to the giant planets), where M is mass. The C-type planetesimals starting between the giant planets are shown as light blue dots, and the outer-disk planetesimals as dark blue dots, initially between 8.0 and 13.0 AU. For all planetesimals, filled dots are used if they are inside the main asteroid belt and smaller open dots otherwise. The approximate boundaries of the main belt are drawn with dashed curves. The bottom panel combines the end state of the giant planet migration simulation (including only those planetesimals that finish in the asteroid belt) with the results of simulations of inner disk material (semimajor axis $a < 2$) evolved for 150 Myr (see Fig. 4), reproducing successful terrestrial planet simulations[8].

and planetesimals (small), while the planetesimal population exterior to Jupiter is partitioned between inter-planetary belts and a trans-Neptunian disk (8–13 AU). The planetesimals from the inner disk are considered to be 'S type' and those from the outer regions 'C type'. The computation of gas drag assumes 100-km-diameter planetesimals and uses a radial gas density profile taken directly from hydrodynamic simulations[4] (see Supplementary Information for details).

The inward migration of the giant planets shepherds much of the S-type material inward by resonant trapping, eccentricity excitation and gas drag. The mass of the disk inside 1 AU doubles, reaching $\sim 2M_\oplus$. This reshaped inner disk constitutes the initial condition for terrestrial planet formation. However, a fraction of the inner disk (~14%) is scattered outward, ending up beyond 3 AU. During the subsequent outward migration of the giant planets, this scattered disk of S-type material is encountered again. Of this material, a small fraction (~0.5%) is scattered inward and left decoupled from Jupiter in the asteroid belt region as the planets migrate away. The giant planets then encounter the material in the Jupiter–Neptune formation region, some of which (~0.5%) is also scattered into the asteroid belt. Finally, the giant planets encounter the disk of material beyond Neptune (within 13 AU) of which only ~0.025% reaches a final orbit in the asteroid belt. When the giant planets have finished their migration, the asteroid belt population is in place, whereas the terrestrial planets require an additional ~30 Myr to complete their accretion.

The asteroid belt implanted in the simulations is composed of two separate populations: the S-type bodies originally from within 3.0 AU, and the C types from between the giant planets and from 8.0 to 13.0 AU. The present-day asteroid belt consists of more than just S- and C-type asteroids, but this diversity is expected to result from compositional gradients within each parent population (Supplementary Information). There is a correlation between the initial and final locations of implanted asteroids (Fig. 3a). Thus, S-type objects dominate in the inner belt, while C-type objects dominate in the outer belt (Fig. 3b). Both types of asteroid share similar distributions of eccentricity and inclination (Fig. 3c, d). The present-day asteroid belt is expected to have had its eccentricities and inclinations reshuffled during the so-called late heavy bombardment (LHB)[13,14]; the final orbital distribution in our simulations matches the conditions required by LHB models.

Given the overall efficiency of implantation of ~0.07%, our model yields $\sim 1.3 \times 10^{-3}\ M_\oplus$ of S-type asteroids at the time of the dissipation of the solar nebula. In the subsequent 4.5 Gyr, this population will be depleted by 50–90% during the LHB event[13,14] and by a further factor of ~2–3 by chaotic diffusion[15]. The present-day asteroid belt is estimated to have a mass of $6 \times 10^{-4}\ M_\oplus$, of which 1/4 is S-type and 3/4 is C-type[12]. Thus our result is consistent within a factor of a few with the S-type portion of the asteroid belt.

The C-type share of the asteroid belt is determined by the total mass of planetesimals between the giant planets and between 8 and 13 AU, which are not known a priori. Requiring that the mass of implanted C-type material be three times that of the S-type, and given the implantation efficiencies reported above, this implies that the following

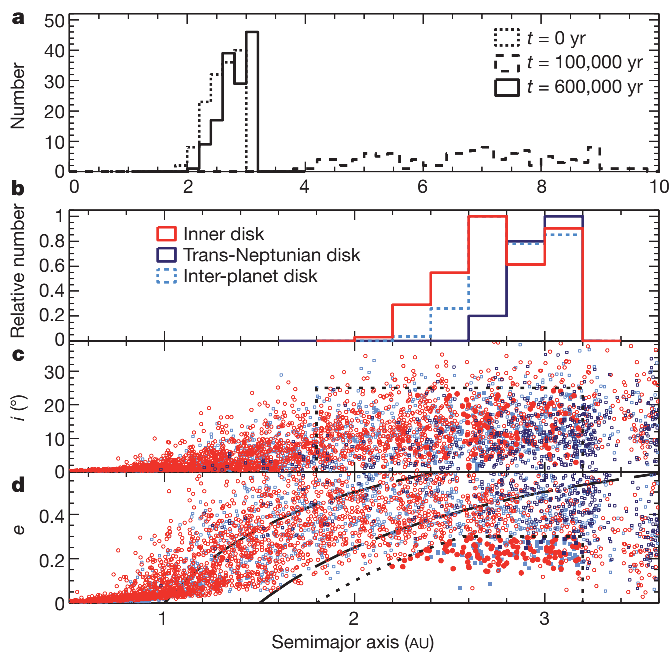


Figure 3 | Distributions of 100-km planetesimals at the end of giant planet migration. **a**, The semimajor axis distribution for the bodies of the inner disk that are implanted in the asteroid belt are plotted at three times: the beginning of the simulation (dotted histogram), at the end of inward planet migration (dashed) and at the end of outward migration (solid). There is a tendency for S-type planetesimals to be implanted near their original location. Thus, the outer edge of their final distribution is related to the original outer edge of the S-type disk, which in turn is related to the initial location of Jupiter. **b**, The final relative numbers of the S-type (red histogram), the inter-planet population (light blue) and the outer-disk (dark blue) planetesimals that are implanted in the asteroid belt are shown as a function of semimajor axis. The orbital inclination (**c**) and eccentricity (**d**) are plotted as a function of semimajor axis, with the same symbols used in Fig. 2. The dotted lines show the extent of the asteroid belt region for both inclination and eccentricity, and the dashed lines show the limits for perihelion less than 1.0 (left line) and 1.5 (right line). Most of the outer-disk material on planet-crossing orbits has high eccentricity, while many of the objects from between the giant planets were scattered earlier and therefore damped to lower-eccentricity planet-crossing orbits.

amount of material is left over from the giant planet accretion process: $\sim 0.8\ M_{\oplus}$ of material between the giant planets; $\sim 16\ M_{\oplus}$ of planetesimals from the 8.0–13 AU region; or some combination of the two.

Our simulations also found C-type material placed onto orbits crossing the still-forming terrestrial planets. For every C-type planetesimal from beyond 8 AU that was implanted in the outer asteroid belt, 11–28 C-type planetesimals ended up on high-eccentricity orbits that enter the terrestrial-planet-forming region (with perihelion $q < 1.0$–1.5 AU; see Fig. 3), and may represent a source of water for Earth[16]. For the Jupiter–Uranus region this ratio is 15–20, and for the Uranus–Neptune region it is 8–15. Thus, depending on which region dominated the implantation of C-type asteroids, we expect that $(3-11) \times 10^{-2}\ M_{\oplus}$ of C-type material entered the terrestrial planet region. This exceeds by a factor of 6–22 the minimal mass required to bring the current amount of water to the Earth ($\sim 5 \times 10^{-4}\ M_{\oplus}$; ref. 17), assuming that C-type planetesimals are 10% water by mass[18].

We now consider the terrestrial planets. The migration of Jupiter creates a truncated inner disk matching initial conditions of previously successful simulations of terrestrial planet formation[8], though there is a slight build-up of dynamically excited planetary embryos at 1.0 AU. Thus, we ran simulations of the accretion of the surviving objects for 150 Myr. Earth and Venus grow within the 0.7–1 AU annulus, accreting most of the mass, while Mars is formed from embryos scattered out beyond the edge of the truncated disk. Our final distribution of planet mass versus distance quantitatively reproduces the large mass ratio existing between Earth and Mars, and also matches quantitative metrics of orbital excitation (Fig. 4).

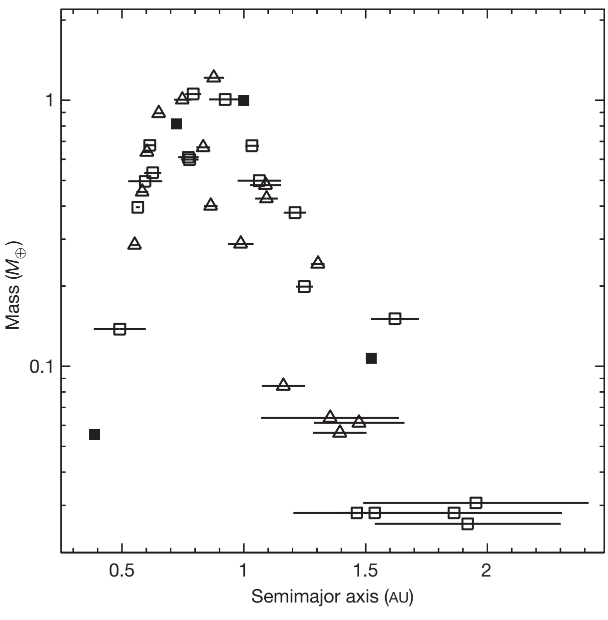


Figure 4 | Results of the eight terrestrial planet simulations. The mass versus semimajor axis of the synthetic planets (open symbols) is compared to the real planets (filled squares). The triangles refer to simulations starting with 40 embryos of $\sim 0.051\ M_{\oplus}$, and squares to simulations from 80 embryos of $\sim 0.026\ M_{\oplus}$. The horizontal error bars show the perihelion–aphelion excursion of each planet along their orbits. The initial planetesimal disk had an inner edge at 0.7 AU to replicate previous work[8], and an outer edge at ~ 1.0 AU owing to the truncation caused by the inward and outward migration of the giant planets as described in the text. Half of the original mass of the disk interior to Jupiter ($1.85 M_{\oplus}$) was in ~ 727 planetesimals. At the end of giant planet migration, the evolution of all objects inward of 2 AU was continued for 150 Myr, still accounting for the influence from Jupiter and Saturn. Collisions of embryos with each other and with planetesimals were assumed fully accretional. For this set of eight simulations, the average normalized angular momentum deficit[30] was 0.0011 ± 0.0006, as compared to 0.0018 for the current Solar System. Similarly, the radial mass concentration[30] was 83.8 ± 12.8 as compared to 89.9 for the current Solar System.

Similar qualitative and quantitative results were found for a number of migration schemes, a range of migration and gas disk dissipation timescales, and a range of gas density and planetesimal sizes (all described in Supplementary Information). Our results represent a major shift in the understanding of the early evolution of the inner Solar System. In our scheme, C-type asteroids form between and beyond the giant planets, nearer to comets than to S-type asteroids. This could explain the substantial physical differences between S-type and C-type asteroids, and also the physical similarities between the latter and comets (as shown by the Stardust mission and micrometeorite samples[19,20]; see Supplementary Information for more on physical properties).

If Jupiter and Saturn have migrated substantially, then their birth region could have been closer to the estimated location of the snow line (the expected condensation front for water) at ~ 3 AU (ref. 21), rather than out beyond 5 AU. Also, substantial migration is a point of similarity with observed extrasolar planetary systems, in which migration is seemingly ubiquitous—extrasolar giant planets are commonly found at ~ 1.5 AU (refs 2, 22). However, a difference between our Solar System and the currently known extrasolar systems is that, according to our results, Jupiter 'tacked' at 1.5 AU and then migrated outward, owing to the presence of Saturn.

Received 1 September 2010; accepted 1 April 2011.
Published online 5 June 2011.

1. Haisch, K. E. Jr, Lada, E. A. & Lada, C. J. Disk frequencies and lifetimes in young clusters. *Astrophys. J.* **553,** L153–L156 (2001).
2. Armitage, P. J. Massive planet migration: theoretical predictions and comparison with observations. *Astrophys. J.* **665,** 1381–1390 (2007).
3. Masset, F. & Snellgrove, M. Reversing type II migration: resonance trapping of a lighter giant protoplanet. *Mon. Not. R. Astron. Soc.* **320,** L55–L59 (2001).
4. Morbidelli, A. & Crida, A. The dynamics of Jupiter and Saturn in the gaseous protoplanetary disk. *Icarus* **191,** 158–171 (2007).
5. Pierens, A. & Nelson, R. P. Constraints on resonant-trapping for two planets embedded in a protoplanetary disc. *Astron. Astrophys.* **482,** 333–340 (2008).
6. Kleine, T. *et al.* Hf-W chronology of the accretion and early evolution of asteroids and terrestrial planets. *Geochim. Cosmochim. Acta* **73,** 5150–5188 (2009).
7. Wetherill, G. W. in *Protostars and Planets* (ed. Gehrels, T.) 565–598 (IAU Colloquium 52, International Astronomical Union, 1978).
8. Hansen, B. M. S. Formation of the terrestrial planets from a narrow annulus. *Astrophys. J.* **703,** 1131–1140 (2009).
9. Lin, D. N. C. & Papaloizou, J. On the tidal interaction between protoplanets and the protoplanetary disk. III — Orbital migration of protoplanets. *Astrophys. J.* **309,** 846–857 (1986).
10. Crida, A. Minimum mass solar nebulae and planetary migration. *Astrophys. J.* **698,** 606–614 (2009).
11. Gradie, J. & Tedesco, E. Compositional structure of the asteroid belt. *Science* **216,** 1405–1407 (1982).
12. Mothé-Diniz, T., Carvano, J. M., Á. & Lazzaro, D. Distribution of taxonomic classes in the main belt of asteroids. *Icarus* **162,** 10–21 (2003).
13. Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435,** 466–469 (2005).
14. Morbidelli, A., Brasser, R., Gomes, R., Levison, H. F. & Tsiganis, K. Evidence from the asteroid belt for a violent past evolution of Jupiter's orbit. *Astron. J.* **140,** 1391–1401 (2010).
15. Minton, D. A. & Malhotra, R. Dynamical erosion of the asteroid belt and implications for large impacts in the inner Solar System. *Icarus* **207,** 744–757 (2010).
16. Morbidelli, A. *et al.* Source regions and time scales for the delivery of water to Earth. *Meteorit. Planet. Sci.* **35,** 1309–1320 (2000).
17. Lécuyer, M., Gillet, P. & Robert, F. The hydrogen isotope composition of sea water and the global water cycle. *Chem. Geol.* **145,** 249–261 (1998).
18. Abe, Y., Ohtani, E., Okuchi, T., Righter, K. & Drake, M. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 413–433 (Univ. Arizona Press, 2000).
19. Brownlee, D. *et al.* Comet 81P/Wild 2 under a microscope. *Science* **314,** 1711–1716 (2006).
20. Gounelle, M. *et al.* in *The Solar System Beyond Neptune* (eds Barucci, M. A. *et al.*) 525–541 (Univ. Arizona Press, 2008).
21. Ciesla, F. J. & Cuzzi, J. N. The evolution of the water distribution in a viscous protoplanetary disk. *Icarus* **181,** 178–204 (2006).
22. Butler, R. P. *et al.* Catalog of nearby exoplanets. *Astrophys. J.* **646,** 505–522 (2006).
23. Paardekooper, S. & Mellema, G. Halting type I planet migration in non-isothermal disks. *Astron. Astrophys.* **459,** L17–L20 (2006).
24. Paardekooper, S. & Papaloizou, J. C. B. On disc protoplanet interactions in a non-barotropic disc with thermal diffusion. *Astron. Astrophys.* **485,** 877–895 (2008).
25. Kley, W. & Crida, A. Migration of protoplanets in radiative discs. *Astron. Astrophys.* **487,** L9–L12 (2008).

26. Masset, F. S. & Casoli, J. On the horseshoe drag of a low-mass planet. II. Migration in adiabatic disks. *Astrophys. J.* **703,** 857–876 (2009).
27. Masset, F. S. & Papaloizou, J. C. B. Runaway migration and the formation of hot Jupiters. *Astrophys. J.* **588,** 494–508 (2003).
28. Morbidelli, A., Tsiganis, K., Crida, A., Levison, H. F. & Gomes, R. Dynamics of the giant planets of the Solar System in the gaseous protoplanetary disk and their relationship to the current orbital architecture. *Astron. J.* **134,** 1790–1798 (2007).
29. Batygin, K. & Brown, M. E. Early dynamical evolution of the Solar System: pinning down the initial conditions of the Nice model. *Astrophys. J.* **716,** 1323–1331 (2010).
30. Raymond, S. N., O'Brien, D. P., Morbidelli, A. & Kaib, N. A. Building the terrestrial planets: constrained accretion in the inner Solar System. *Icarus* **203,** 644–662 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements K.J.W. and A.M. were supported by the Helmholtz Alliances 'Planetary Evolution and Life' programme. S.N.R and A.M.M. were supported by the EPOV and PNP programmes of CNRS. D.P.O'B. was supported by the NASA PG&G programme. A.M.M. was also supported by the NASA post-doctoral programme and the Goddard Center for Astrobiology. We thank the Isaac Newton Institute DDP programme for hosting some of us at the initial stage of the project; we also thank J. Chambers for comments that improved the text. Computations were done on the CRIMSON Beowulf cluster at OCA.

Author Contributions K.J.W. managed the simulations and analysis and was the primary writer of the manuscript. A.M. initiated the project, updated and tested software, ran and analysed simulations, and wrote significant parts of the manuscript. S.N.R. helped initiate the project, advised on simulations and contributed substantially to the manuscript. D.P.O'B. helped initiate the project and assisted in writing. A.M.M. assisted in software updates and in writing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.J.W. (kwalsh@boulder.swri.edu).